CHAPTER
TWELVE

OPTOELECTRONIC DEVICES

12.1 THE P-N JUNCTION DIODE

This Chapter is devoted to the wide variety of optoelectronic devices that are made in semiconductor materials. Many (although not all) involve the absorption and emission of light, and most are based around the surprisingly versatile structure mentioned at the end of Chapter 11, the **p-n junction diode**. We shall begin by considering the electronic aspects of this device.

In its simplest form, a p-n junction diode can be constructed as a **homojunction**, using an abrupt change between p- and n-type regions in a single piece of semiconductor. Figure 12.1-1 illustrates the formation of a diode, in a one-dimensional 'thought experiment' based on the ideas of the previous Chapter. The p- and n-regions might initially be considered to be separated slabs of material. When these are placed in intimate contact, the difference in hole concentration between the p- and n-type regions should result in the diffusion of holes from the former into the latter (Figure 12.1-1a). In the same way, electrons ought to diffuse in the reverse direction, from the n-type material into the p-type. If the two carrier types meet near the interface, they will recombine. The effect should then be that a layer - known as the **depletion region** - will be almost entirely denuded of carriers (Figure 12.1-1b).

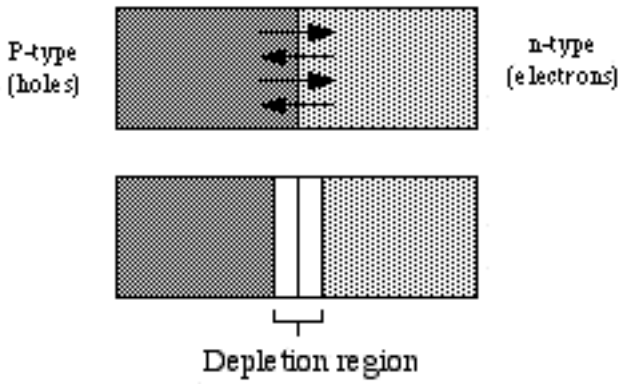


Figure 12.1-1 Formation of a p-n junction between p- and n-type material: a) counterdiffusion of majority carriers, and b) the fully formed junction.

A logical question to ask is, what determines the width of the depletion region? For example, is it possible that the counter-diffusion process would continue indefinitely, until all the free carriers have been used up, so that the depletion region extends right through both layers of material? The short answer is that the process is self-limiting. As we shall see, carrier recombination in the depletion region sets up a space charge; this in turn induces an electric field, which acts as an obstacle to further motion of carriers, so that the depletion region quickly stabilises as a thin layer. However, in practise, the dynamics of junction formation are irrelevant to the operation of the device, and we need only consider the situation after equilibrium has been established.

THE DISTRIBUTION OF CHARGE AND ELECTRIC FIELD INSIDE THE DEPLETION LAYER

Due to the lack of free carriers, the depletion region must be locally charged, since the charges of the dopant ions that are embedded in the lattice are no longer compensated by those of the mobile carriers. The sign of this 'exposed charge' varies. On the p-side, for

example, it is negative, due to exposure of the charge of the ionised acceptors. In contrast, the charging is positive on the n-side, because the positive charges of donor ions are now exposed. To a first approximation, the charge distribution within each region is uniform (although we shall review this later). Thus, assuming total ionization of the dopant atoms, a space extending from (say) $x = -x_{dp}$ to $x = 0$ in the p-type material may be assumed to have a local charge density $\rho = -eN_a$, while in the n-type material the space between $x = 0$ and $x = x_{dn}$ has a charge density eN_d (Figure 12.1-2a). Here x_{dn} and x_{dp} are the lengths of the p- and n-sides of the depletion layer; these are related to the total width w of the region by $w = x_{dp} + x_{dn}$.

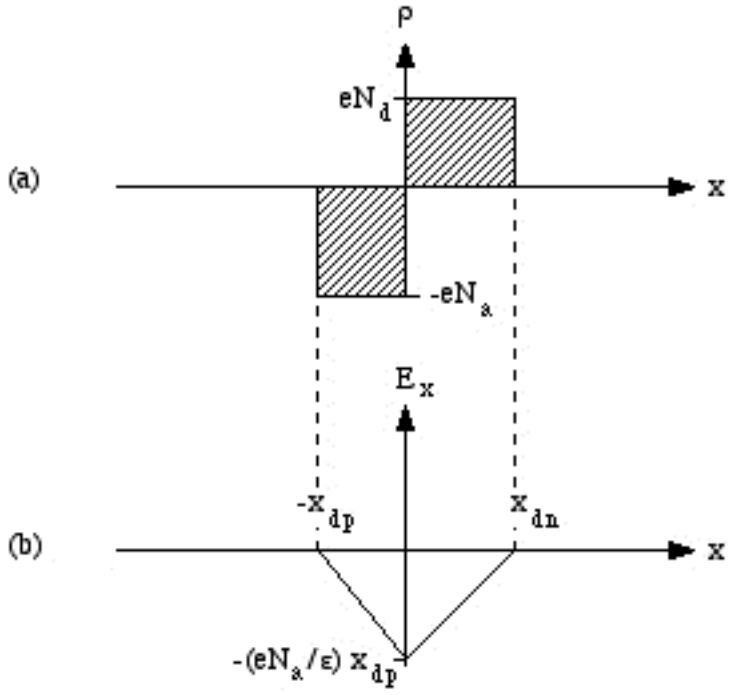


Figure 12.1-2 a) charge density and b) electric field, for a p-n junction in equilibrium.

Now, when taken as a whole, the complete structure must be electrically neutral, so the exposed charges must balance each other. We may therefore put:

$$x_{dp}N_a = x_{dn}N_d \qquad 12.1\text{-}1$$

This allows us to write the width of the depletion layer as:

$$w = x_{dn}\{1 + N_d/N_a\} \qquad 12.1\text{-}2$$

Note that the 'split' of the depletion layer between the p- and n-sides may be highly unequal, if $N_a \neq N_d$. For example, the p-side might be much more heavily doped than the n-side; in this case (known as a **p^+-n junction**), we find that $x_{dp} \ll x_{dn}$, so the depletion layer extends further into the n-side.

We can use Figure 12.1-2a to find the electric field resulting from the charge distribution. The calculation is based Gauss' law, which requires that the electric flux flowing out of a closed surface equals the charge contained inside it. Taking the enclosed volume to be a parallel-sided slab of unit cross-sectional area lying between two planes at $-x_{dp}$ and x, the enclosed charge is given by:

$$q = -eN_a\,(x + x_{dp}) \qquad (\text{for } -x_{dp} \le x \le 0)$$
$$= -eN_d\,(x_{dn} - x) \qquad (\text{for } 0 \le x \le x_{dn})$$

12.1-3

This must give rise to an equal flux density, which (by symmetry) has only an x-component D_x. As a result, the corresponding electric field $E_x = D_x/\varepsilon$ is given by:

$$E_x = -(eN_a/\varepsilon)\,(x + x_{dp}) \qquad (\text{for } -x_{dp} \le x \le 0)$$
$$= -(eN_d/\varepsilon)\,(x_{dn} - x) \qquad (\text{for } 0 \le x \le x_{dn})$$

12.1-4

where ε is the dielectric constant of the semiconductor. This electric field variation is shown plotted in Figure 12.1-2b. As can be seen, the field varies linearly on each side of the depletion region. On the p-side, it falls from $E_x = 0$ at $x = -x_{dp}$ to $E_x = -(eN_a/\varepsilon)\,x_{dp}$ at $x = 0$; similarly, on the n-side, it rises from this minimum value to zero at $x = x_{dn}$. Now, the existence of an electric field within the depletion region suggests that there will be a voltage between different points. We might therefore find the p.d. between (for example) the left-hand edge of the depletion region and an arbitrary point x by direct integration, i.e. as:

$$V(x) = \int_{-x_{dp}}^{x} -E_x\,dx$$

12.1-5

Substituting from Equation 12.1-4, we then obtain:

$$V(x) = (eN_a/\varepsilon)\,\{x^2/2 + xx_{dp} + x_{dp}^{\,2}/2\} \qquad (\text{for } -x_{dp} \le x \le 0)$$
$$= (eN_a/\varepsilon)\,x_{dp}^{\,2}/2 + (eN_d/\varepsilon)\,\{xx_{dn} - x^2/2\} \qquad (\text{for } 0 \le x \le x_{dn})$$

12.1-6

This distribution is shown schematically in Figure 12.1-3. Throughout the depletion region, V(x) varies parabolically, following a smooth S-shaped curve from zero at $-x_{dp}$ to a maximum (which we shall refer to as the **built-in voltage**, V_{bi}) at x_{dp}. Since the electric field is zero at either end of the depletion layer, dV/dx is also zero at these points.

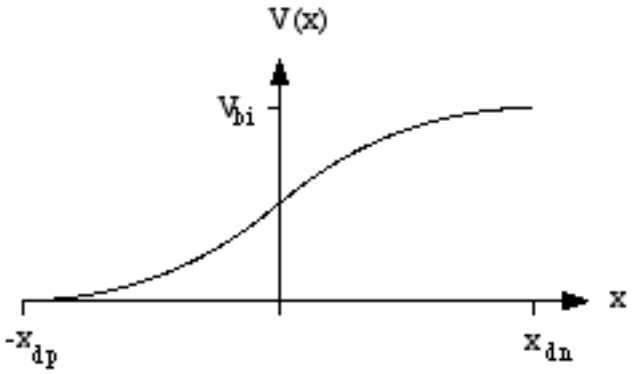


Figure 12.1-3 Potential distribution inside a p-n junction at equilibrium.

Clearly, the built-in voltage may be found as $V_{bi} = V(x_{dn})$; substituting into Equation 12.1-6, we get:

$$V_{bi} = (eN_a/\varepsilon)\,x_{dp}^{\,2}/2 + (eN_d/\varepsilon)\,x_{dn}^{\,2}/2$$

12.1-7

Substituting for x_{dp} from Equation 12.1-1, and for x_{dn} from Equation 12.1-2, we can find V_{bi} in terms of the depletion layer width w as:

$$V_{bi} = (ew^2/2\varepsilon)\,\{N_aN_d / (N_a + N_d)\}$$

12.1-8

Conversely, we can find the width w in terms of the built-in voltage by inverting Equation 12.1-8, as:

$$w = \sqrt{\{(2\varepsilon V_{bi}/e)\ (N_a + N_d) / N_a N_d\}} \qquad 12.1\text{-}9$$

At this point, we have found a closed-form expression for the depletion layer width, which contains just one unknown parameter, $V_{bi.}$ We shall now consider the factors that determine $V_{bi.}$ itself.

THE BAND DIAGRAM FOR A P-N JUNCTION

We start by returning to the results of Chapter 11, where we drew energy band diagrams for isolated regions of n- and p-type material. How should these be modified for the composite p-n junction? We would expect some features - the relative spacings of the conduction band, the Fermi level and the valence band, for example - to be preserved at a reasonable distance from the depletion layer. Inside the depletion region, the band structure will clearly be modified. However, we may make one deduction: the Fermi level (for which the occupation probability is 0.5) must be the same in all regions, because this level is derived purely from considerations of thermal equilibrium. Armed with these two rules, we can guess that the band diagram for the p-n junction must be constructed from the two diagrams of Figure 11.6-2 as shown in Figure 12.1-4.

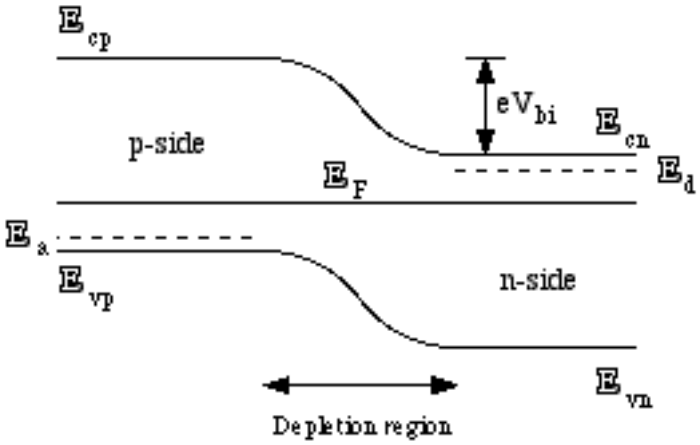


Figure 12.1-4 Energy band structure for a p-n junction in equilibrium.

We start by considering the situation outside the depletion region. Here the bands are flat, and the main features are spaced as before, but the absolute values of the energy levels in the two halves of the structure are shifted to make the Fermi levels coincide. After this shift, the conduction band level is clearly different in the p- and n-regions. We will define the values of $\mathbb{E}_c$ in these two regions as as $\mathbb{E}_{cp}$ and $\mathbb{E}_{cn}$. Now, remember that the position of the Fermi level in n-type material was previously defined relative to the conduction band, in terms of the density of donor atoms. Using Equation 11.7-13, we may therefore define the Fermi level $\mathbb{E}_{Fn}$ as:

$$\mathbb{E}_{Fn} \approx \mathbb{E}_{cn} - kT \log_e(N_c/N_d) \qquad 12.1\text{-}10$$

Similarly, the position of the Fermi level in p-type material can be defined relative to the valence band in terms of the acceptor density. However, it will be more convenient to relate $\mathbb{E}_{Fp}$ to $\mathbb{E}_{cp}$ instead. We can do this by combining Equations 11.7-12 and 11.7-14, which gives:

$$\mathbb{E}_{Fp} \approx \mathbb{E}_{cp} - kT \log_e(N_a N_c / n_i^2)$$

12.1-11

We now have two results for the Fermi level, one for each side of the junction. According to our previous argument, they should be equal. Equating them allows us to find the difference between the conduction band levels on the p- and n-sides, as:

$$\mathbb{E}_{cp} - \mathbb{E}_{cn} = kT \log_e\{N_a N_d / n_i^2\}$$

12.1-12

Now, a difference in the energy of a conduction electron on the two halves of the junction represents a potential barrier, which acts to prevent the transit of electrons across the depletion layer. There is a similar barrier to hole motion, which can be calculated from the difference between the valence band levels on the p- and n-sides. Overall, the energy barrier is equivalent to a built-in voltage across the structure of $V_{bi} = \{\mathbb{E}_{cp} - \mathbb{E}_{cn}\}/e$. V_{bi} is therefore given by:

$$V_{bi} = (kT/e) \log_e\{N_a N_d / n_i^2\}$$

12.1-13

Equation 12.1-13 is the missing link, which will allow us to calculate the width of the depletion layer.

DESIGN EXAMPLE

We can find the parameters of a typical p-n junction as follows. Assuming that the material is silicon (for which $n_i \approx 1.4 \times 10^{16}$ m^{-3} at T = 300 K), and that the donor and acceptor densities in the p- and n-regions have the typical values of $N_a = 10^{24}$ m^{-3}, $N_d = 10^{22}$ m^{-3}, respectively (so this is a p^+-n junction), Equation 12.1-13 implies that the built-in voltage is $V_{bi} = \{1.38 \times 10^{-23} \times 300 / 1.6 \times 10^{-19}\} \log_e\{10^{24} \times 10^{22} / 2 \times 10^{32}\} = 0.82$ V.

The depletion layer width is then given by Equation 12.1-9 (assuming that $\varepsilon = 11.7\ \varepsilon_0$ for Si) as $w = \sqrt{\{(2 \times 11.7 \times 8.85 \times 10^{-12} \times 0.82 / 1.6 \times 10^{-19})\ (10^{24} + 10^{22}) / (10^{24} \times 10^{22})\}}$ m = 0.33 μm. Most of this layer extends into the n-region, since $N_a \gg N_d$. The salient features of a p-n junction are therefore a built-in voltage of the order of a volt, dropped across a layer less than a micron wide. This is equivalent to an electric field of $\approx 2.5 \times 10^6$ V/m, a high value.

THE BAND DIAGRAM INSIDE THE DEPLETION REGION

We may exploit the relation between voltage and the energy of a conduction electron once again, to fill in the missing features of the band diagram inside the depletion region. We have already seen that the voltage V(x) relative to the left-hand edge of the depletion layer varies parabolically. Consequently, the conduction band energy level will also be spatially varying, following:

$$\mathbb{E}_c(x) = \mathbb{E}_{cp} - eV(x)$$

12.1-14

The conduction band therefore bends parabolically across the depletion layer, as in Figure 12.1-4. The valence band varies similarly, maintaining a constant energy gap $\mathbb{E}_g$ between the two levels.

Knowledge of this band structure may be used to calculate the free carrier distribution throughout the depletion layer. For example, if we substitute Equation 12.1-14 into Equation 11.7-4, we can find the variation of the conduction electron density as:

$$n(x) = N_c \exp\{-[\mathbb{E}_c(x) - \mathbb{E}_F]/kT\} \qquad 12.1\text{-}15$$

A typical variation of n(x) is shown in Figure 12.1-5. Note that n(x) does not abruptly fall to zero at the edge of the depletion layer, as we have assumed so far; instead, the variation is smooth. However, the drop from the equilibrium value in the n-type material outside the depletion layer (labelled nn) is sufficiently rapid for our assumed charge distribution (Figure 12.1-2a) to be a good approximation to the truth.

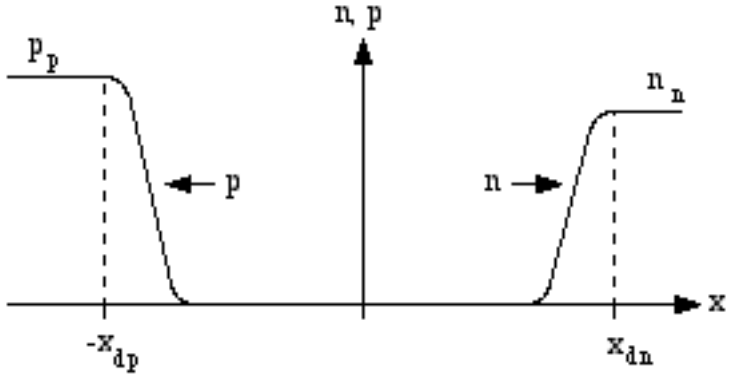


Figure 12.1-5 Carrier distribution for a p-n junction in equilibrium.

Our model is therefore reasonably self-consistent. The corresponding hole density variation p(x) is also shown. The equilibrium hole density p_p in the p-type material may be vastly different from n_n if the junction is asymmetric (i.e., if $N_a \neq N_d$).

Note also that the value of n at the depletion layer edge in the p-region (which we might call n_p) is not identically zero, although it will generally be fairly small. From Equation 12.1-15, we can see that $n_n = n(x_{dn}) = \exp\{-(\mathbb{E}_{cn} - \mathbb{E}_F)/kT\}$, so that $n_p = n(-x_{dp})$ is given by:

$$n_p = n_n \exp(-eV_{bi}/kT) \qquad 12.1\text{-}16$$

The value of p at the depletion layer edge in the n-region (termed pn) is found in a similar way as:

$$p_n = p_p \exp(-eV_{bi}/kT) \qquad 12.1\text{-}17$$

where $p_p = \exp\{-(\mathbb{E}_F - \mathbb{E}_{vp})/kT\}$, and $\mathbb{E}_{vp}$ is the valence band level in the p-region (Figure 12.1-4). Given that V_{bi} is significantly greater than 25 mV (i.e. kT/e at 300 K), we find that n_p, $p_n \ll n_n$, p_p.

We can use these distributions to examine the details of the equilibrium inside the depletion layer, for example for electron motion. In Chapter 11, we saw that the current density for electrons is composed of two components (due to drift and diffusion, respectively). If the two contributions are calculated, it can be shown that they balance exactly at any point inside the depletion layer, and that the net current density is zero (although each component may individually be very large).

A similar dynamic balance may be found for holes, so that equilibrium is established for both types of charge carrier. This contrasts with the situation outside the depletion layer;

here there is no field, and no concentration gradient either, so that the two current components are zero independently for both species.

THE BAND DIAGRAM OF A BIASSED P-N JUNCTION

Suppose that electrical contacts are now attached to the two ends of the diode, so that external voltages can be applied. We shall assume that these are metal, and of a particular type that forms an **ohmic contact** with the semiconductor. In this case, the metal can supply or receive any number of electrons, so the carrier density at the interface will be the equilibrium value. For example, aluminium forms an ohmic contact on p- or heavily doped n-type Si. Ohmic contacts may be also made to intrinsic or n-type Si, via an intermediate layer of n^+-Si. However, other combinations (e.g., Al on n^-- GaAs) form a rectifying contact called a **Schottky barrier**, which is itself a form of diode.

Since the depletion layer has a very low carrier concentration, we would expect the bulk of any applied voltage to be dropped across this highly resistive layer. If we make the n-type side more negative than the p-type by V volts - a condition known as **forward bias** - this will tend to reduce the built-in barrier from eV_{bi} to $e\{V_{bi} - V\}$. As a result, the band diagram will alter as shown in Figure 12.1-6

On the other hand, if the n-side is made more positive than the p-side - when the junction is said to be in **reverse bias** - the applied voltage will tend to increase the size of the built-in potential barrier. In this case, the band diagram will modify as shown in Figure 12.1-7.

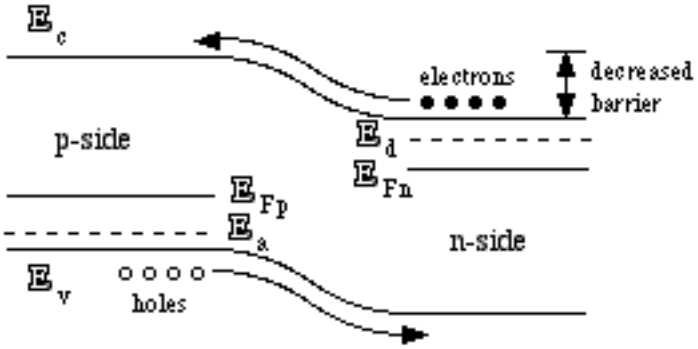


Figure 12.1-6 Energy band structure of a p-n junction in forward bias.

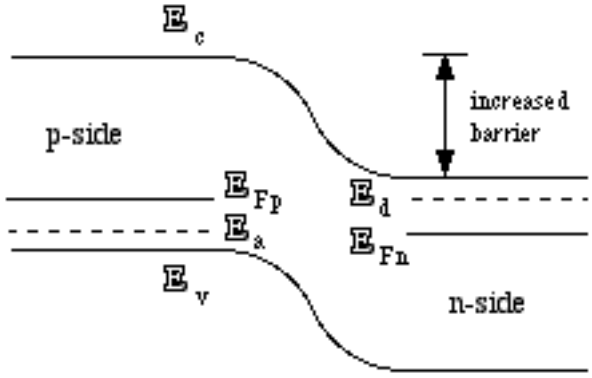


Figure 12.1-7 Energy band structure of a p-n junction in reverse bias.

CURRENT-VOLTAGE CHARACTERISTICS

The reduction in the barrier height in a forward-biassed junction will naturally make it easier for electrons to travel across the depletion region from the n-side to the p-side, and for holes to move in the reverse direction. The arrival of these carriers at the depletion layer edges will then cause a change in the minority carrier densities. The new values n_p' and p_n' may be found by substituting $V_{bi} - V$ for V_{bi} in Equations 12.1-16 and 12.1-17. If this is done, it is easy to see that:

$$n_p' = n_p \exp(eV/kT) \quad \text{and} \quad p_n' = p_n \exp(eV/kT) \qquad 12.1\text{-}18$$

Clearly, n_p' will be much greater than the equilibrium value n_p (and similarly $p_n' \gg p_n$) for values of V greater than about 25 mV. Even a small forward bias will therefore produce a considerable increase in minority carrier density near the depletion layer edges. This is called **minority carrier injection**. Now, any such increase in np (for example) must follow from an actual transport of electrons from the n- to the p-side. However, it is important to recognise that the densities of the majority carriers are generally so much greater than those of the minority carriers that the fractional change in n_n is very small. Its value (and that of p_p) may thus be considered unchanged to a good approximation.

The fate of the injected carriers depends on the geometry. If the diode is short enough, recombination may be neglected, and the carriers simply diffuse to the contacts. In this case, we must return to the results of Chapter 11, where we showed that the equations governing carrier diffusion had the form $\partial n,p/\partial t = D_{e,h}\partial^2 n,p/\partial x^2$. In the steady-state, $\partial n,p/\partial t = 0$, so the carrier densities must vary linearly with distance. The solution is then as shown in Figure 12.1-8; on the p-side, the electron concentration falls linearly from n_p' at the depletion layer edge ($x = -x_{dp}$) to the equilibrium value n_p at the contact ($x = -x_p$). Similarly, on the n-side, the hole concentration falls from p_n' to p_n between $x = x_{dn}$ and $x = x_n$.

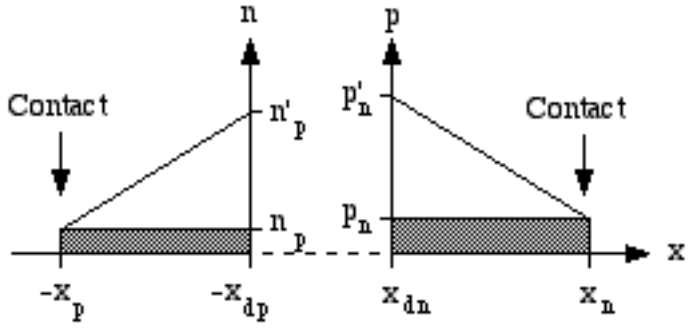


Figure 12.1-8 Minority carrier distribution outside the depletion layer of a forward biased p-n junction.

Despite the absence of appreciable fields outside the depletion region, some current must still flow to the contacts, because the concentration gradients give rise to diffusion currents. The total current density is the sum of contributions from electron and hole diffusion, namely:

$$J = D_e e\, dn/dx - D_h e\, dp/dx = D_e e\{n_p' - n_p\}/L_p + D_h e\{p_n' - p_n\}/L_n \qquad 12.1\text{-}19$$

where the lengths L_p and L_n are given by $L_p = x_p - x_{dp}$ and $L_n = x_n - x_{dn}$. Note that if the doping is unequal, the two components in Equation 12.1-19 may differ vastly in size. For example, in a p^+-n junction, the majority hole density p_p is much greater than the

corresponding electron density n_n, so the minority hole densities p_n' and p_n are greater than $n_p'$and n_p. The current is then carried mainly by holes. Substituting for n_p' and p_n', we find the following relation between J and V:

$$J = J_s \{\exp(eV/kT) - 1\} \qquad 12.1\text{-}20$$

where J_s (known as the **reverse saturation current density**) is given by:

$$J_s = D_e e n_p/L_p + D_h e p_n/L_n \qquad 12.1\text{-}21$$

This is the well-known exponential characteristic of a diode, shown in Figure 12.1-9. For a device of cross-sectional area A, the current is simply I = J x A, so the I - V characteristic has a similar shape.

When V is zero, the current is also zero. For a small positive voltage, however, the forward current can be extremely high. In contrast, the reverse current rapidly saturates at $I_s = J_s$ x A (which is typically small) when V is negative. As a result, the diode acts as a rectifier, presenting an extremely high impedance to reverse voltages and a low impedance to forward ones. Note that the behaviour of a real diode departs from the ideal at high reverse voltages, a feature we will return to shortly.

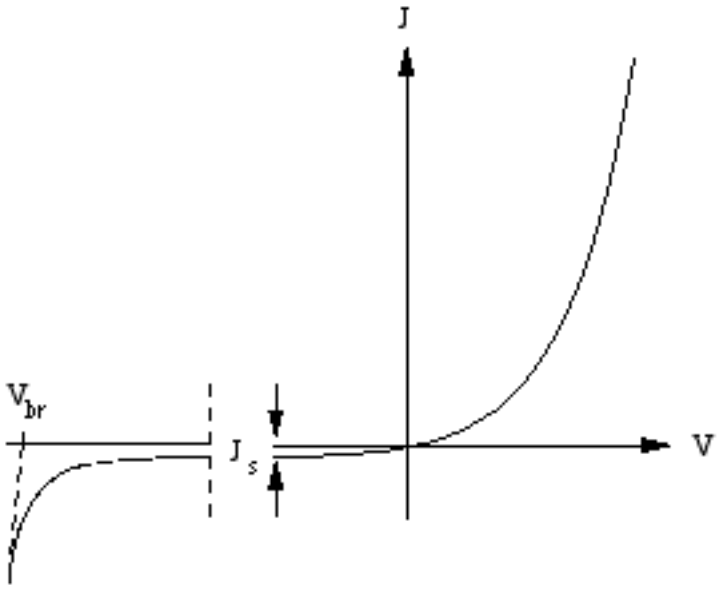


Figure 12.1-9 J-V characteristics for a p-n junction diode.

THE EFFECT OF RECOMBINATION

We now examine how this characteristic is modified, if the distance to the contacts is large enough for recombination to be important (which is the case in light-emitting devices). Once again, we start by returning to the results of Chapter 11. In the steady state, the electron density (for example) on the p-side is governed by:

$$D_e \partial^2 n/\partial x^2 - (n - n_p)/\tau_e = 0 \qquad 12.1\text{-}22$$

where we have written n_p for the equilibrium level n_E. This is a standard second-order differential equation, which can be solved exactly, subject to the boundary conditions of $n = n_p'$ at the edge of the depletion layer ($x = -x_{dp}$) and $n = n_p$ at the contact ($x = -x_p$). Provided x_p is sufficiently large, however, a reasonable approximation to the solution is given by:

$$n(x) - n_p = \{n_p' - n_p\} \exp\{(x + x_{dp}) / \sqrt{(D_e\tau_e)}\}$$

12.1-23

Equation 12.1-23 shows that the excess electron density decays exponentially away from the depletion layer edge, falling to 1/e of its initial value in a length $L_e = \sqrt{(D_e\tau_e)}$. This is known as the **diffusion length** of electrons in the p-type material. The excess hole density on the n-side varies in a similar way, decaying to 1/e of its initial value in a corresponding length $L_h = \sqrt{(D_h\tau_h)}$.

Now, the value of J_e at $x = -x_{dp}$ is the total injected electron current density. Performing the necessary differentiation, we find that:

$$J_e = D_e e\{n_p' - n_p\} / L_e$$

12.1-24

This is qualitatively very similar to the equivalent term in Equation 12.1-19, apart from the replacement of L_p by L_e. Consequently, we may deduce that our later result for the J - V characteristic (Equations 12.1-20 and 12.1-21) will still be valid, merely provided we replace L_p by L_e and L_n by L_h. A diode therefore still behaves as a rectifier, even when recombination is significant.

THE EFFECT OF VOLTAGE ON THE DEPLETION LAYER WIDTH

We now consider how the width of the depletion layer is modified by the application of a voltage. To find the new width w(V), we merely substitute $V_{bi} - V$ for V in Equation 12.1-9, which gives:

$$w(V) = \sqrt{\{1 - V/V_{bi}\}}\, w(0)$$

12.1-25

where w(0) is the width of the unbiassed junction. The effect of a negative value of V is therefore to increase the depletion layer width, while a positive value has the opposite effect. Now, according to Equation 12.1-25, w can become negative for voltages greater than V_{bi}; this is clearly unrealistic, and merely shows that our original approximation concerning the charge distribution inside the depletion layer has broken down at this point.

AVALANCHE BREAKDOWN

At high reverse voltages, our approximations will also break down (although in a slightly different manner). Consider the band diagram shown in Figure 12.1-10. An electron moving to the right will travel down the sloping region of the conduction band in a series of 'hops', with each abrupt decrease in energy corresponding to a collision with the lattice.

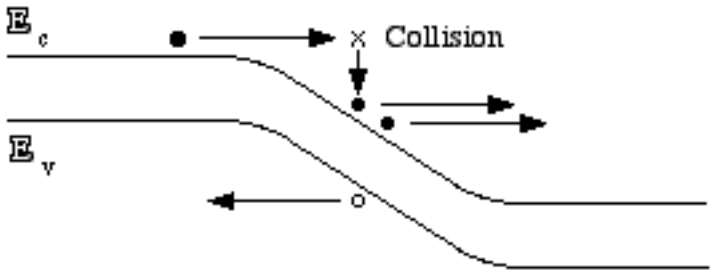


Figure 12.1-10 Carrier multiplication by avalanche breakdown under high reverse bias.

If the electric field is high enough, the electron may gain a considerable amount of kinetic energy between collisions. Consequently, it is possible for the energy given up at a collision to be greater than $\mathbb{E}_g$. In this case, the energy is sufficient to ionise a lattice atom, by promoting a valence electron to the conduction band. If this occurs, the number of carriers will increase, since there will now be two electrons moving to the right, and one hole moving to the left. After further acceleration by the field, the electrons may perform further ionisations, and so on. Under high reverse bias, a shower of new carriers may therefore be created. This leads to a sudden rise in the reverse current - a behaviour known as **avalanche breakdown** - at the characteristic breakdown voltage V_{br} shown in Figure 12.1-9. This effect is exploited in the avalanche photodiode, which we will describe later.

HETEROJUNCTION DIODES

Up until now, we have assumed that the materials on either side of the junction are of essentially the same type, with any differences introduced through doping. This need not be the case - for example, we have already mentioned the metal-semiconductor junction. Alternatively, different semiconductors may be used. Clearly, there are certain constraints; it must be feasible to grow the layered structure as a perfect crystal, which requires the two materials to be lattice-matched. Despite this restriction, it is entirely possible to fabricate diodes with different band-gaps on either side of the junction. In this case, the structure is known as a **heterojunction**.

Heterojunctions are of great importance in optoelectronics, since they allow two important new features to be built in to the basic diode structure. Firstly, the difference in refractive index between the two materials may be used to provide optical confinement, as in the heterostructure waveguides described in Chapter 9. Secondly, the modification to the energy band diagram introduced by the variation in $\mathbb{E}_g$ may be used to provide different potential barriers to the motion of electrons and holes, allowing each type of carrier to be confined independently to a particular region of space. As a result, the performance of light-emitting devices may be improved enormously.

The energy band diagram of a typical heterojunction is shown in Figure 12.1-11, under equilibrium conditions. Here the energy gap $\mathbb{E}_{g1}$ on the p-side is greater than the value $\mathbb{E}_{g2}$ on the n-side, so the structure is known as a P-n anisotype heterojunction (the upper-case letter denoting the material with the larger energy gap). Outside the depletion layer, the diagram is constructed as before, using the continuity of the Fermi level to determine the absolute levels of the conduction and valence bands. Inside the depletion layer itself, however, the construction is slightly different.

The analysis required to determine the band variation may be adapted from that used for the homojunction. The exposed charge distribution of Figure 12.1-2a may again be used; however, the electric field variation shown in Figure 12.1-2b must be modified, to allow for the different dielectric constants ε_1 and ε_2 in the two materials. The former value must be used in the upper line in Equation 12.1-4, the latter in the lower one. Integration of the electric field to find the potential variation V(x) across the depletion layer may then be carried out much as before. If this is done, a slightly different proportion of the built-in voltage V_{bi} is found to be dropped across each side of the junction.

Due to the difference in energy gap between the two sides of the junction, it is no longer possible to construct the conduction and valence bands by assuming that they follow the potential variation V(x). In fact, as we shall see shortly, both bands must be discontinuous. Instead, we must rely on the assumption that the vacuum level - the level above the conduction band - is continuous and follows V(x). Generally, this band is displaced by an

amount χ from the conduction band level E_c, where χ is a quantity known as the **electron affinity** (the energy required to liberate an electron from the conduction band) and ϕ above the Fermi level, where ϕ is the **work function**.

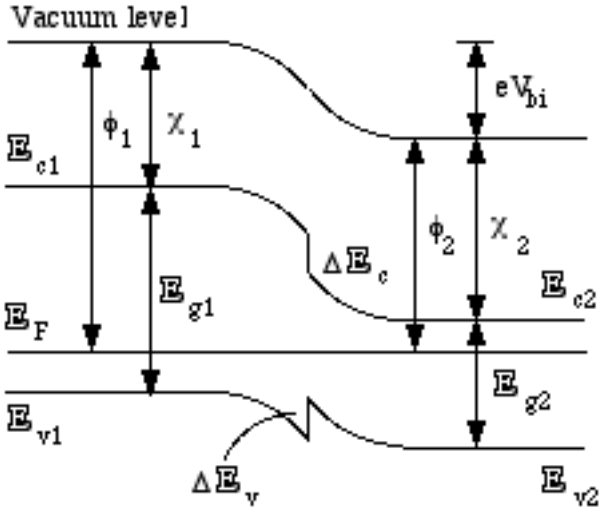


Figure 12.1-11 Energy band structure for a P-n heterojunction in equilibrium.

To complete the diagram, the vacuum level is drawn in on either side of the junction. The difference in height of this level then determines the built-in voltage V_{bi}, allowing the width of the depletion layer to be found as before. Within the depletion layer, the variation of the vacuum level follows V(x). The conduction band is then drawn in at a height χ_1 below the vacuum level on the p-side, and χ_2 below on the n-side, where χ_1 and χ_2 are the electron affinities in the two materials.

This construction leads to an abrupt discontinuity $\Delta E_c \approx \chi_2 - \chi_1$ in the conduction band level at the boundary between the semiconductors. The valence band is then drawn in at a height E_{g1} below the conduction band level on the p-side, and E_{g2} below on the n-side, leading to a discontinuity ΔE_v at the interface. The net result is that the potential barrier faced by electrons has height $eV_{bi} + \Delta E_c$, while that for holes has height $eV_{bi} - \Delta E_v$, making it harder for electrons to cross the junction, and easier for holes. The same state of affairs exists in the P-p isotype heterojunction, formed between two different p-type semiconductors. Other combinations may be envisaged where the reverse is true, so the holes face a higher barrier; this is the case in both p-N and n-N structures.

PRACTICAL P-N JUNCTION DIODES

p-n junction diodes may be fabricated as multilayer structures, for example by growing a layer of p-type material on an n-type substrate using an epitaxial process. In this case, the resulting junction is said to be **abrupt**, and its main features are as described above. Alternatively, diodes may be fabricated by diffusion, for example by heavily doping a region near the surface of an n-type substrate with acceptors. In this case, the properties of the semiconductor do not change rapidly between the p- and n-type regions, but are instead vary more gradually, following the distribution of dopant beneath the surface. In this type of **graded junction**, the charge distribution in the depletion region varies more smoothly than we have assumed, and the electric field distribution must also be modified accordingly. Nonetheless, the operating principles described above are still broadly correct.

12.2 ELECTRO-OPTIC SEMICONDUCTOR DEVICES

We now return to a problem we touched on briefly in Chapters 9 and 10, that of constructing electro- optic waveguide devices in semiconducting materials. Here, the difficulty is that the presence of carriers prevents the development of a high electric field across the guide. The p-n junction provides a solution to the problem. For example, Figure 12.2-1a shows a cross-section of a typical phase modulator, built from an n-type InP homostructure strip-loaded waveguide on a <100> n^+ substrate. The guide is capped with a p^+ layer, thus forming a p^+-n homojunction at the top of the guide, with the bulk of the depletion region located inside the guide itself.

One ohmic contact is placed on top of the p^+ layer, while the other is connected to the substrate. When the junction is reverse biased, the highly resistive depletion region grows downwards, so that all of the applied voltage is eventually dropped across the guide layer. The field is then sufficient to obtain a usable index change through the relatively weak r_{41} electro-optic coefficient, thus inducing an optical phase change. Figure 12.2-1b is an electron microscope photograph of a similar device, based on an InP/InGaAsP heterostructure.

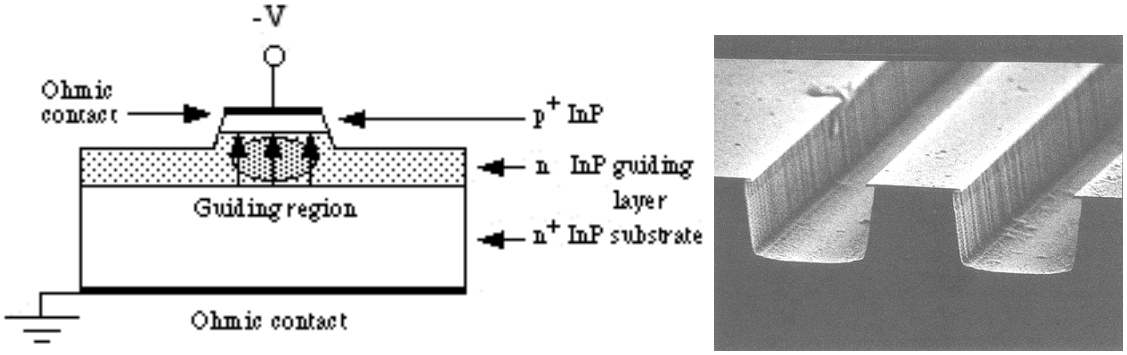


Figure 12.2-1 a) Cross-section of an InP integrated optic phase modulator; b) SEM photograph of an actual device (photograph courtesy P.M.Rogers, British Telecom Research Laboratories).

Electro-optic directional coupler switches may be constructed in a similar way, using two closely- spaced strip-loaded guides. For example, Figure 12.2-2 shows a cross-section through a typical directional coupler, this time constructed using n^--on-n^+ GaAs homostructure guides.

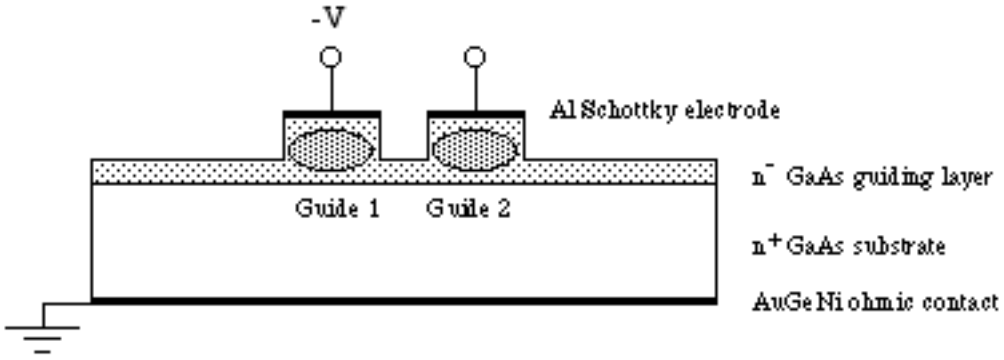


Figure 12.2-2 Cross-section of a GaAs integrated optic directional coupler (after H.Inoue et al., Appl. Opt. 25, 1484-1490 (1986)).

Here, a much simpler electrode structure is used, with Al electrodes deposited directly on top of the two guides. As a result, these contacts are not ohmic (although the substrate contact is), but are instead of the rectifying Schottky barrier type mentioned earlier. These also act as diodes, so that operation in reverse bias is much as described above. To switch the coupler, a negative voltage is applied to just one of the Schottky electrodes, thus inducing an electric field which desynchronises the guides.

Electrically switchable devices may even be constructed using semiconductors with weak or negligible electro-optic coefficients. We already know from Chapters 3 and 9 that the refractive index changes used in homostructure guides are obtained through variations in doping, which cause changes in the free carrier concentration. A similar index change can be created dynamically, by the electrically- controlled injection of minority carriers, this time using a forward-biased p-n junction.

For example, Figure 12.2-3 shows the cross-section of an X-switch formed in silicon (which does not exhibit a linear electro-optic effect at all). Again, a p^+-n junction is constructed at the top of the guide. When the junction is forward-biased, holes are injected into the guiding region, thus generating a refractive index change. Optical absorption is avoided by operating at a wavelength outside that required for band-to-band transitions, e.g. $\lambda_0 = 1.3\ \mu m$. Similar devices may be made in InGaAsP, and the index change caused by carrier injection is actually an order of magnitude larger than that available through the electro-optic effect. However, switching speeds are limited by the lifetime of the injected carriers, which must disappear through recombination when the forward current is switched off.

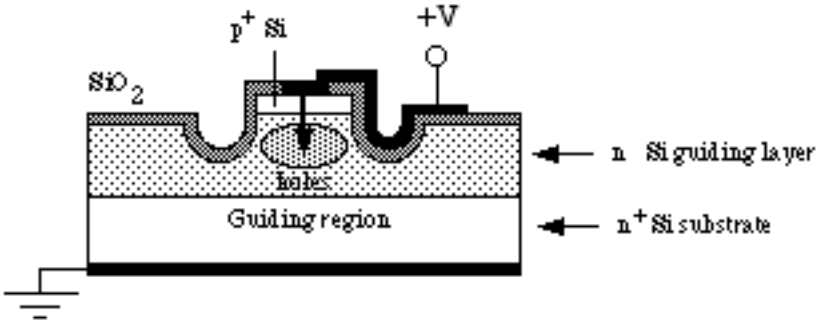


Figure 12.2-3 Cross-section of a Si-based carrier-injection switch (after J.P. Lorenzo et al., Appl. Phys. Lett. 51, 6-8 (1987)).

12.3 PHOTODIODES

We shall now consider the p-n junction in the alternative role of a light detector, or **photodiode**. In Chapter 11, we noted that absorption in a bulk semiconductor can induce band-to-band transitions, causing a measurable increase in conductivity. However, the basic weaknesses of photoconductive detection are a high intrinsic level of noise (because of the presence of thermally excited carriers), and a trade-off between speed and sensitivity (due to recombination of carriers before they can be collected). The solution is to arrange for the absorption to occur in a region devoid of thermal carriers, and then use a high electric field to sweep the photo-generated carriers apart before they can recombine. This can be done inside the depletion layer of a p-n junction. Figure 12.3-1 shows the excitation of an electron from the valence band into the conduction band in this region, by a photon of energy $hf > \mathbb{E}_c - \mathbb{E}_v$. Clearly, the built-in field will tend to drive the photo-generated electron

out of the depletion region towards the n-side, and the corresponding hole towards the p-side.

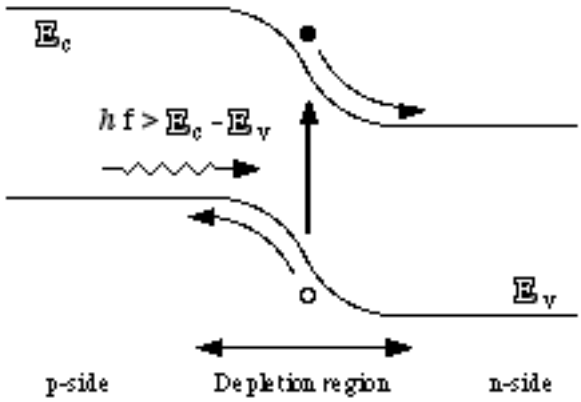


Figure 12.3-1 Generation of an electron-hole pair by a photon inside the depletion region of a p-n junction.

Two types of photodiode are in common use. Most are surface-entry devices, where the light is input through (for example) the p^+ side of a p^+-n junction. Ideally, the light should pass into the depletion layer without suffering significant absorption in the p^+ region, because carrier pairs generated outside the depletion layer will not be separated. Once inside this layer, the absorption should be total, because any light transmitted into the n-region will also be wasted. The depletion layer should therefore be as thick as possible. Side-entry devices, where the light is passed directly into the depletion layer, are another possibility. In this case, the useful absorbing region may be extremely long. However, the depletion layer should again be thick, or the light will spill out on either side.

Once out of the depletion layer, the carriers diffuse to the contacts. As a result, there is an increase in the reverse current, so that the J - V characteristic is modified in the presence of light as shown in Figure 12.3-2. Photodiodes are normally operated at a fixed reverse bias voltage, when they act as a current source. In this case, a reverse current (known as the **dark current**) flows in the absence of illumination and contributes to noise. For a large reverse bias, this current will be Is.

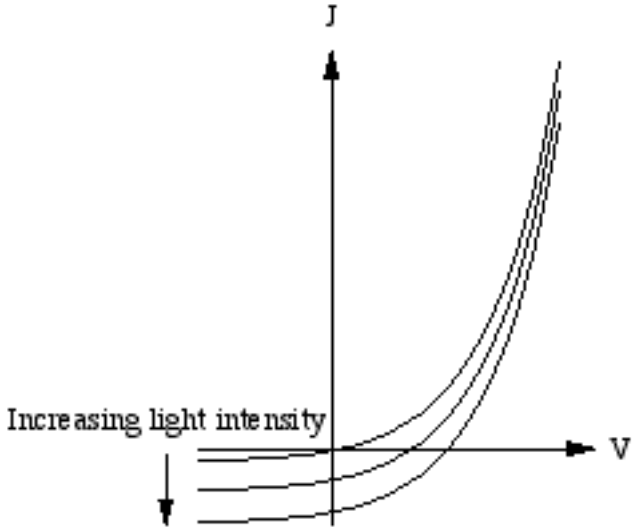


Figure 12.3-2 J-V characteristic for a p-n junction diode in the presence of a light flux.

We may estimate the responsivity of a photodiode as follows. Assuming that the optical power falling on the detecting area is P, the number of photons arriving per second is $P/hf = P\lambda_0/hc$. However, not all of these will be absorbed, and not all the carriers generated by those that are will actually reach the contacts. We may model this by defining a **quantum efficiency** η, representing the number of useful carrier-pairs generated per photon. Clearly, the maximum possible value of η is 100%. In this case, every photon generates one electron (which is collected on the n-side) and one hole (which arrives on the p-side). The net effect is that one unit of charge crosses the device from side-to-side. The total charge crossing in unit time, which is the **photocurrent** I_P, is then:

$$I_P = \eta P e \lambda_0 / hc \qquad 12.3\text{-}1$$

so the photocurrent is linearly proportional to the optical power, a most useful result which holds good over a wide dynamic range. The **responsivity** $R = I_P/P$ is therefore given by:

$$R = \eta e \lambda_0 / hc \qquad 12.3\text{-}2$$

Assuming uniform quantum efficiency, the responsivity should then be linearly proportional to the wavelength λ_0. However, this is only the case for photons with sufficient energy to promote an interband transition. Less energetic photons will not contribute to the photocurrent at all, so the responsivity must fall to zero for $\lambda_g > hc/\mathbb{E}_g$ (or $\lambda_g > 1.24 / e\mathbb{E}_g$, where λ_g is in μm and $\mathbb{E}_g$ is in eV).

DESIGN EXAMPLE

We may calculate the theoretical responsivity of a silicon p-n junction photodiode as follows. $\mathbb{E}_g$ for Si is 1.14 eV at 300 K, so the optimum wavelength is $\lambda_g = 1.24 / 1.14 = 1.09$ μm. Assuming a uniform quantum efficiency of 100%, $R_{max} = 1.6 \times 10^{-19} \times 1.09 \times 10^{-6} / (6.62 \times 10^{-34} \times 3 \times 10^{8}) = 0.88$ A/W. The theoretical responsivity is then as shown by the dashed line in Figure 12.3-3.

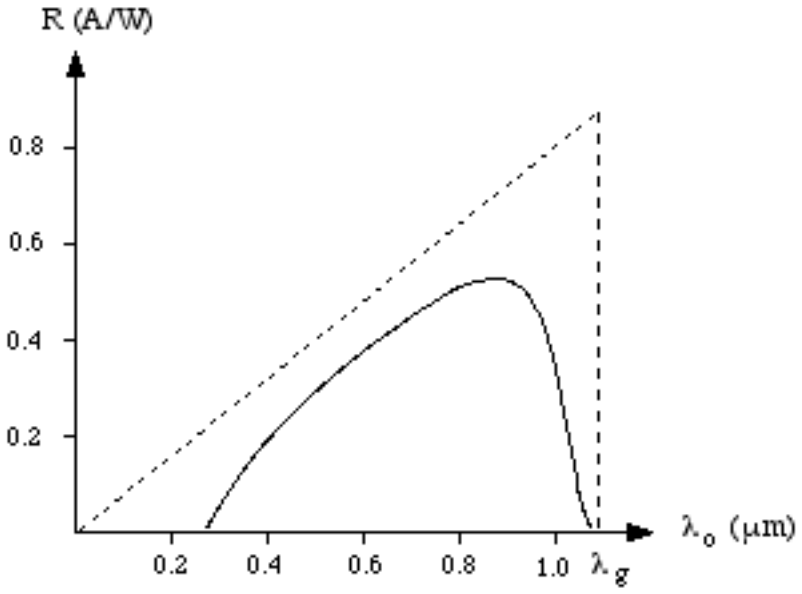


Figure 12.3-3 Ideal (dashed line) and actual (full line) responsivity of a silicon p-n junction photodiode.

In practice, the quantum efficiency of a surface-entry device is variable, and peaks at about 80%. At short wavelengths, η is reduced, because the attenuation is so strong that the light is absorbed before it reaches the depletion layer, and at long wavelengths it falls because the light is transmitted right through this layer. Furthermore, since Si has an indirect band-gap, η is reduced further near λ_g, because a simultaneous interaction with a high-energy phonon is required for momentum conservation in the band-to-band transition, a relatively unlikely event. The characteristics of a real Si photodiode are therefore as given by the full line in Figure 12.3-3, which shows a useful responsivity only up to $\lambda_0 \approx 0.85$ μm. For high sensitivity at the near-infrared wavelengths used in second-generation fibre communication systems, materials with smaller bandgaps are required. Ge (useful up to $\lambda_0 \approx 1.55$ μm) or GaInAs (up to 1.67 μm) are both suitable.

PIN PHOTODIODES

The basic limitation of a p-n junction photodiode - that the depletion layer is so thin that radiation of long wavelength is only weakly absorbed - is overcome in the p-i-n structure shown in Figure 12.3-4a (known as a **PIN photodiode**). Here, a region of intrinsic or lightly doped material is introduced between two heavily doped p- and n-type regions. Because the doping is so low in this region, the depletion layer can then be arranged to extend right through it under a modest reverse bias, yielding the electric field variation of Figure 12.3-4b. The effective depletion layer width may therefore be fixed at a value far greater than the 'natural' one, approximately the width of the intrinsic region.

Figure 12.3-4c shows a cross-section of a typical GaInAs PIN photodiode. The n- region is a mesa of $Ga_{0.47}In_{0.53}As$ (a particular composition that is lattice-matched to InP). This is grown epitaxially on an n^+ InP substrate, and the p^+ region is then made by zinc diffusion. The structure is therefore a hybrid, part homo- and part hetero-junction. In this device, light enters through the InP substrate, which has an energy gap of 1.35 eV and so is transparent to wavelengths longer than about 0.92 μm. This avoids the absorption of any light at the important 1.3 and 1.55 μm wavelengths outside the depletion layer. The quantum efficiency is then almost uniform between 1.0 and 1.6 μm, just below λ_g for GaInAs.

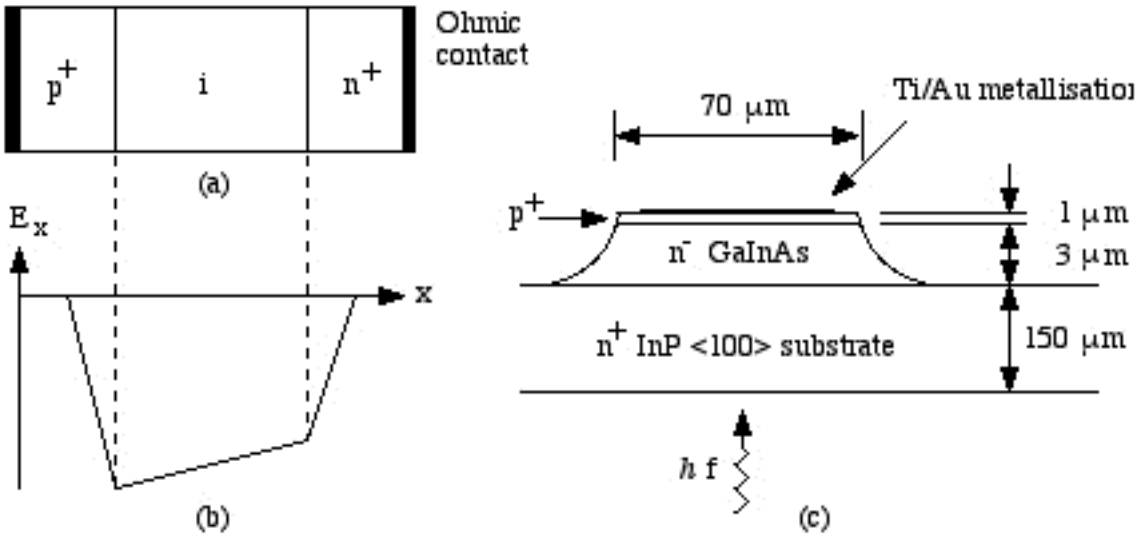


Figure 12.3-4 a) A p-i-n structure, b) the corresponding electric field variation, and c) a substrate- entry GaInAs PIN photodiode (after R.H. Moss Br. Telecom J. 1, 7-22 (1983)).

AVALANCHE PHOTODIODES

It is possible to make use of the high-field characteristics of the p-n junction to construct a photodiode with internal gain. In the **avalanche photodiode** (or APD), impact ionisation is used as a method of multiplying the photocurrent before it enters the circuit of any electrical amplifier, thus providing an increase in receiver sensitivity. Figure 12.3-5 shows the reach-through avalanche photodiode (RAPD), which illustrates the basic principle involved.

The RAPD is a four-layer device, with a p^+-i-p-n^+ structure (Figure 12.3-5a). Under reverse bias, most of the applied voltage is dropped across the p-n^+ junction, although the exact fraction depends on the doping levels. As the voltage is increased, the depletion layer associated with this junction widens, until it just reaches through to the intrinsic region. The electric field distribution is then as shown in Figure 12.3-5b. In the depletion region, the field is extremely high; normally, the peak field is held at ≈ 10% below the value at which avalanche breakdown occurs. In the long intrinsic region, however, the electric field is only moderate. Light enters the structure through the p^+ layer, and travels through to the intrinsic region, where it is absorbed. The photogenerated carriers drift in the moderate electric field, with electrons travelling towards the p-n^+ junction. Carrier multiplication then takes place inside the high-field region by impact ionization. As a result, there is a large increase in the photocurrent, up to 1000 times, which provides the desired improvement in detector sensitivity. However, since the avalanche process is a statistical one, there is a corresponding increase in the noise level.

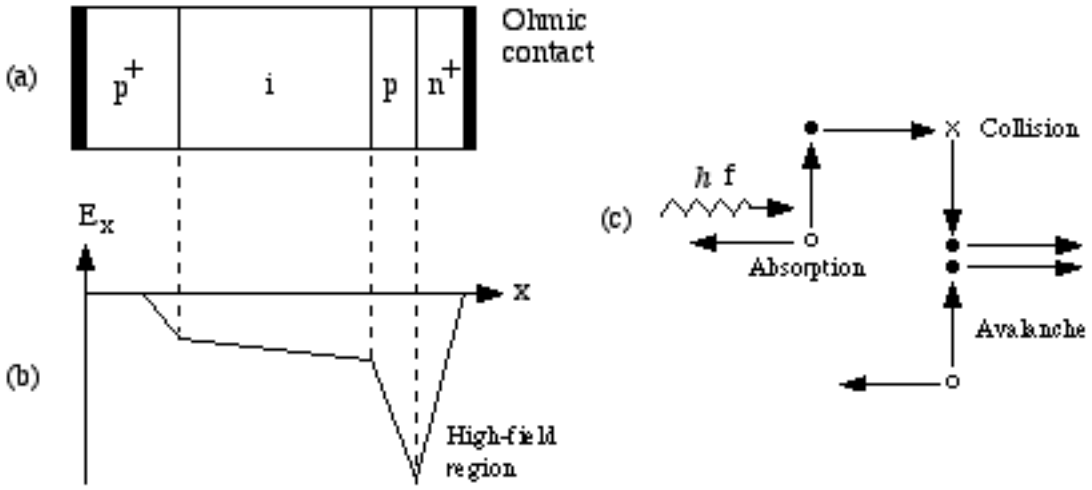


Figure 12.3-5 a) Avalanche photodiode structure, b) the corresponding electric field variation, and c) the processes of absorption and avalanche gain.

WAVEGUIDE-INTEGRATED PHOTODIODES

The integration of photodetectors with waveguides poses an inherent compatibility problem - the former are intended to act as efficient, localised absorbers of light, while the latter should (ideally) be lossless. A number of strategies may be used to circumvent the difficulty. Firstly, the waveguide and the detector may be constructed on different levels of a multilayer, allowing the properties of each level to be tailored individually. Light may then be transferred between the two at the appropriate point, either by directional coupling or by radiation. In the latter case, a standard component such as a grating coupler may be used, or the guide may be induced to cut off. For example, Figure 12.3-6 shows the cross-section of a structure combining a passive guide made in two glassy layers with a p-n junction photodiode formed by a p-type diffusion into an n-type silicon substrate. To couple light

into the detector, the height of the SiO_2 buffer layer is gradually reduced until the high-index substrate approaches the guide sufficiently to induce cutoff. The light is then radiated onto the detector.

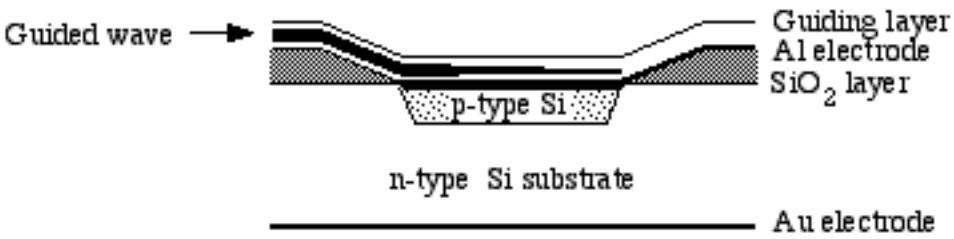


Figure 12.3-6 An Si p-n junction integrated with an overlay guide (after S. Ura et al., IEEE J. Lightwave Tech. LT-4, 913-918 (1986); © IEEE 1986).

Alternatively, waveguides and photodetectors may be constructed on the same level of an integrated structure, and butt-coupled together. Clearly, this requires material properties to be tailored locally. For example, Figure 12.3-7 shows an InP/InGaAsP double-heterostructure channel guide integrated with an GaInAs PIN photodiode. Here, the waveguide is formed first, by conventional methods. A trench with dimensions ≈ 10 μm x 20 μm is then etched right through the guide to the n^+-InP substrate. This trench is filled with n- InGaAs by liquid phase epitaxy (LPE) to form the intrinsic region of the diode, and the upper layer is then Zn-doped to form the p^+ region. The net result is that the channel guide is coupled directly into the depletion region of a vertically oriented PIN photodiode.

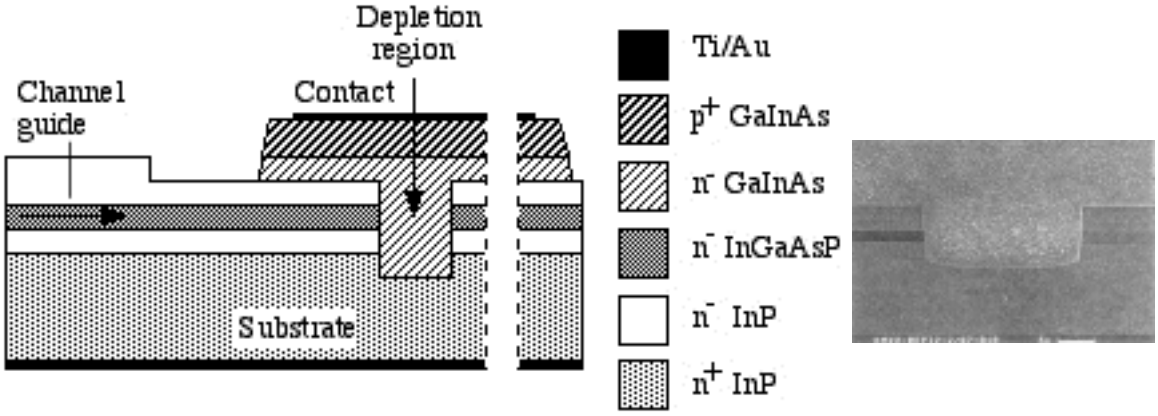


Figure 12.3-7 a) Cross-section of an InGaAs PIN photodiode integrated with an InGaAsP waveguide (after W.Döldissen et al., Elect. Lett. 25, 35-37 (1989)), and b) SEM view in the region of LPE overgrowth (photograph courtesy W. Döldissen, Heinrich-Hertz-Institut).

12.4 THE LIGHT-EMITTING DIODE

We now consider another application of the p-n junction, as a light source. In this incarnation, it is known as a **light-emitting diode** (or LED). Now, in Chapter 11, we described electroluminescence as a mechanism for generating photons in a direct-gap semiconductor. However, an isolated lump of material will not emit significant quantities of light, because in thermal equilibrium at room temperature the number of downward electron transitions is extremely small. To improve the optical output, we must move the material far from equilibrium, so that the rate of spontaneous emission is considerably increased. This might be done by taking (for example) a p-type material, which already contains a large hole density, and pouring electrons into it. This can be done in a forward-biased p-n junction.

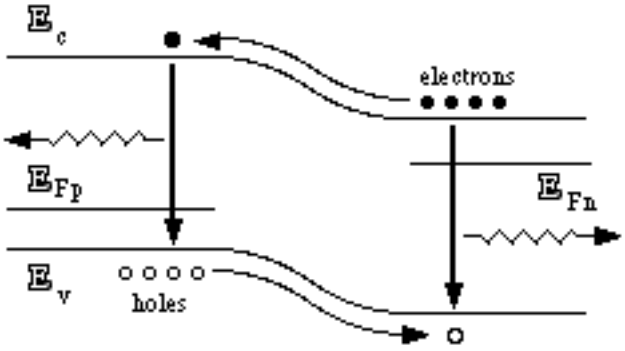


Figure 12.4-1 Spontaneous emission by recombination of injected minority carriers in a forward biased p-n junction.

Figure 12.4-1 illustrates the process; electrons are injected into the p-type material, where they combine with the holes already present. At the same time, holes are injected into the n-type material; however, the junction is normally highly asymmetric, so that light emission takes place mainly from one side.

THE EMISSION SPECTRUM OF AN LED

Once again, we recall that spontaneous emission generates photons that travel in random directions, so the emission is isotropic; as we will show later, this greatly reduces the external efficiency of an LED. The emission is also unpolarized. Furthermore, we note that the output does not consist only of light at the wavelength $\lambda_g = hc/\mathbb{E}_g$, as a two-state model would imply. Because of the wide distribution of carrier energies within the conduction and valence band, the output is incoherent, with a spectrum consisting of a broad range of wavelengths (Figure 12.4-2). The electron distribution normally peaks near an energy ≈ 1/2 kT above the edge of the conduction band, and extends for several kT (say, ≈ 5/2 kT) above $\mathbb{E}_c$. Similarly, the hole distribution peaks at around 1/2 kT below the valence band edge, extending for about 5/2 kT below $\mathbb{E}_v$. Consequently, the possible photon energies lie in the range $\mathbb{E}_g$ to $\mathbb{E}_g$ + 5kT, and the most likely energy is ≈ $\mathbb{E}_g$ + kT.

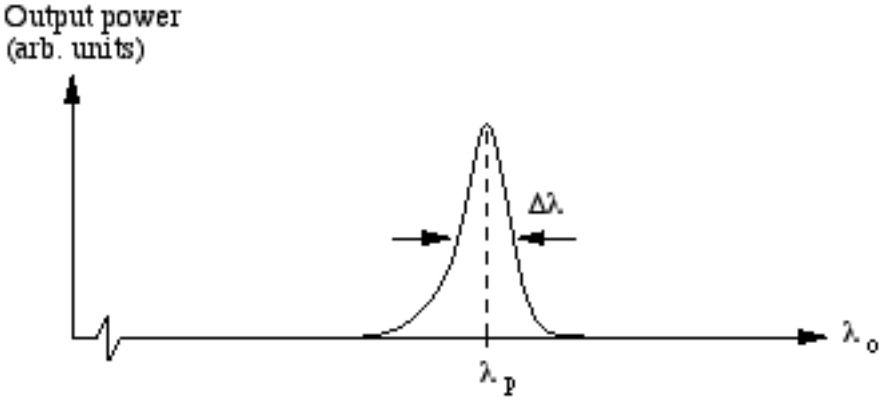


Figure 12.4-2 Representative emission spectrum of an LED.

The optical wavelength corresponding to an energy $\mathbb{E}_g + \alpha kT$ (where α is an arbitrary constant) is $\lambda_0 = hc/\{\mathbb{E}_g + \alpha kT\}$. Since kT is typically much less than $\mathbb{E}_g$, we may put:

$$\lambda_0 \approx (hc/\mathbb{E}_g)\ \{1 - \alpha kT/\mathbb{E}_g\} \qquad 12.4\text{-}1$$

If kT and $\mathbb{E}_g$ are measured in eV, and λ_0 in μm, Equation 12.4-1 may be written in the alternative form $\lambda_0 \approx (1.24/\mathbb{E}_g)\ \{1 - \alpha kT/\mathbb{E}_g\}$. The output from an LED therefore peaks at a wavelength λ_p, given by:

$$\lambda_p \approx (1.24/\mathbb{E}_g)\ \{1 - kT/\mathbb{E}_g\} \qquad 12.4\text{-}2$$

Similarly, the half-power bandwidth of the emission spectrum is $\Delta\lambda$, where:

$$\Delta\lambda \approx 3.1\ kT/\mathbb{E}_g^{\,2} \qquad 12.4\text{-}3$$

Now, $\mathbb{E}_g$ for GaAs is 1.42 eV. For a GaAs LED at room temperature, for example, we therefore find that $\lambda_p \approx (1.24/1.42)\{1 - 0.0258/1.42\} = 0.857$ μm. Similarly, $\Delta\lambda = 3.1 \times 0.0258 / 1.422 = 0.04$ μm, or 40 nm. For materials with smaller bandgaps, designed to emit at longer wavelengths, $\Delta\lambda$ is even larger. For example, an InGaAsP LED operating near $\lambda_0 = 1.3$ μm would have a spectral half-width approaching 100 nm. Consequently, the LED cannot be used as a source for high-speed, long-distance optical communications; the dispersion caused by such an extended spectrum would be far too large. LED transmitters are therefore restricted to short-haul applications.

THE FREQUENCY RESPONSE OF AN LED

We now consider the efficiency of the electroluminescence process inside an LED, which we might term the **internal efficiency** η. As it turns out, η is frequency-dependent. We can illustrate this by calculating the optical output of an LED for a time-varying drive current, assuming a two-state model for simplicity. We start by returning the rate equations of Chapter 11. Considering spontaneous emission to be the dominant process, and the injected carriers to be electrons, the first relevant equation is Equation 11.9-19. Since the system will be significantly perturbed, however, we must replace terms of the form Δn by n, and so on. Furthermore, for simplicity, we shall consider the LED to be a lumped-element device, consisting of a uniform recombination region of cross-sectional area A and depth Δx (i.e., a finite volume v = AΔx). In this case, we may rewrite Equation 11.9-19 as:

$$dn/dt + n/\tau_e = J/e\Delta x = I/ev \qquad 12.4\text{-}4$$

where I = J A is the total current injected into the recombination region. Now, in the general case, the current may consists of a DC bias, with a superimposed AC modulation. We might therefore put:

$$I = I' + I'' \exp(j\omega t) \qquad 12.4\text{-}5$$

Where I' is the DC bias, and I" and ω are the amplitude and angular frequency of the AC modulation, respectively. We assume that I' > I", so that the LED is always under forward bias. We may take the solution for the electron density in a similar form, as:

$$n = n' + n'' \exp(j\omega t) \qquad 12.4\text{-}6$$

Substituting these expressions into Equation 12.4-4, we get:

$$n' + n''(1 + j\omega\tau_e) \exp(j\omega t) = \{I' + I'' \exp(j\omega t)\}\ \tau_e/ev \qquad 12.4\text{-}7$$

Equating the DC and AC terms in Equation 12.4-7, it is clear that we must have:

$$n' = I'\tau_e/ev \quad \text{and} \quad n'' = (I''/ev)\,\{\tau_e / (1 + j\omega\tau_e)\}$$

12.4-8

We now return to the corresponding photon rate equation. For spontaneous emission due to electron injection, the rate of change of photons may be taken as:

$$d\phi/dt = n/\tau_{rr} - \phi/\tau_p$$

12.4-9

Here we have included an additional term $-\phi/\tau_p$, which describes rate of loss of photons from the active volume (since it is now finite). This must be proportional to ϕ itself, and the proportionality constant τp is known as the **photon lifetime**. For an active volume whose longest dimension is L, τ_p is of the order of L/v_g, the time taken for a photon to escape from the region. For typical device dimensions of 100 μm, τ_p is $\approx 10^{-12}$ seconds. Since $1/\tau_p$ is then so large, we may neglect $d\phi/dt$ by comparison, and write:

$$\phi/\tau_p = n/\tau_{rr}$$

12.4-10

For a recombination volume v, the total number of photons escaping per second is $v\phi/\tau_p$. This must represent the external photon flux Φ generated by the diode. Since each photon carries energy $hc/\lambda_g \approx e\mathbb{E}_g$, the optical power P emitted is:

$$P \approx hc\Phi/\lambda g \approx (hc/\lambda_g)\, vn/\tau_{rr}$$

12.4-11

We now assume a solution for P in the form of a DC emission plus an AC modulation term, i.e. as:

$$P = P' + P'' \exp(j\omega t)$$

12.4-12

Substituting this expression into Equation 12.4-11, together with our previous solution for n, we get:

$$P' + P'' \exp(j\omega t) = (hc/\lambda_g)\,[(I'\tau_e/\tau_{rr}e) + (I''/\tau_{rr}e)\,\{\tau_e / (1 + j\omega\tau_e)\}\exp(j\omega t)]$$

12.4-13

Equating the DC and AC components separately in Equation 12.4-13, we then obtain:

$$P' = \eta' hcI'/e\lambda_g \quad \text{and} \quad P'' = \eta'' hcI''/e\lambda_g$$

12.4-14

where we have defined two new quantities, the DC and AC internal efficiencies η' and η'', as:

$$\eta' = \tau_e/\tau_{rr} \quad \text{and} \quad \eta'' = \tau_e/\{\tau_{rr}\,(1 + j\omega\tau_e)\}$$

12.4-15

Equation 12.4-14 shows that the optical output from an LED is linearly proportional to current, for both DC and AC injection. However, the internal efficiency in each case is slightly different. We shall now examine the implications of this result, by considering the likely values of η' and η''.

DC INTERNAL EFFICIENCY

Bearing in mind that $1/\tau_e = 1/\tau_{rr} + 1/\tau_{nr}$, the DC efficiency η' has the value:

$$\eta' = r_{rr} / \{r_{rr} + r_{nr}\}$$

12.4-16

where $r_{rr} = 1/\tau_{rr}$ and $r_{nr} = 1/\tau_{nr}$ are the rates of radiative and non-radiative transitions, respectively. Consequently, the DC efficiency depends on the fraction of all downward transitions that are radiative, a physically convincing result. In simple LEDs, η' may be as high as 50%. Substituting for the constants, the DC output power may be written in the simple form $P' = 1.24\ \eta' I/\lambda_g$, where λ_g is measured in μm.

For a GaAs LED emitting at $\lambda_g \approx 850$ nm, with $\eta' \approx 50\%$, roughly 0.73 mW of optical power should thus be generated per mA of drive current. However, this figure is lowered by absorption, and by total-internal-reflection at the surfaces of the device, as we shall show later.

AC INTERNAL EFFICIENCY

Combining Equations 12.4-15, we note that the AC internal efficiency may also be written as:

$$\eta'' = \eta' / (1 + j\omega\eta'\tau_{rr})\} \qquad 12.4\text{-}17$$

Since η'' is complex, it is clear that the optical modulation cannot be in-phase with the drive current modulation, but must instead lag behind. The magnitude of η'' is given by:

$$|\eta''| = \eta' / \sqrt{\{1 + \omega^2\eta'^2\tau_{rr}^2\}} \qquad 12.4\text{-}18$$

At low frequencies, the AC internal efficiency tends to the DC value, but for angular frequencies greater than the break-point $\omega_c = 1/\eta'\tau_{rr}$, $|\eta''|$ is approximately:

$$|\eta''| \approx 1/\omega\tau_{rr} \qquad 12.4\text{-}19$$

so that $|\eta''|$ is independent of η' at high frequencies.

Figure 12.4-3 shows a logarithmic plot of $|\eta''|$ versus ω, for three values of η'. As can be seen, there is a trade-off between the DC efficiency and the position of the break point.

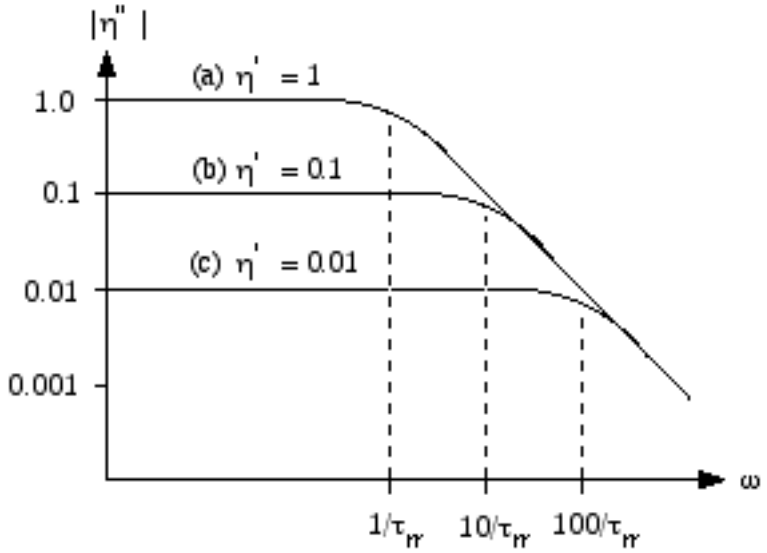


Figure 12.4-3 Variation of the AC internal efficiency η'' of an LED, for three values of DC internal efficiency.

This illustrates a basic limitation of LEDs: they may be either fast, but inefficient, or slow but efficient. However, it is not possible to obtain the best of both worlds. DC efficiency may be optimised by maximising η', by heavily doping the recombination region. The frequency response is clearly limited by τ_{rr}, and in typical devices the break-point occurs at $f_c = \omega_c/2\pi \approx 10$ - 20 MHz. LEDs may therefore only be used as sources for optical links of moderate bit-rate.

THE EXTERNAL EFFICIENCY OF AN LED

Despite the reasonable internal efficiency of an LED, it is an unfortunate truth that very little of this theoretical output is available for use. To illustrate this point we show in Figure 12.4-4a a typical surface-emitting LED; this is a layered p-n structure, in which the optical output is taken from the upper surface. Because the light generation is isotropic, half of the light is emitted downwards, and has no chance of escaping from this surface. Of the half that is emitted upwards, only a fraction escapes, due to the small critical angle at the semiconductor/air interface.

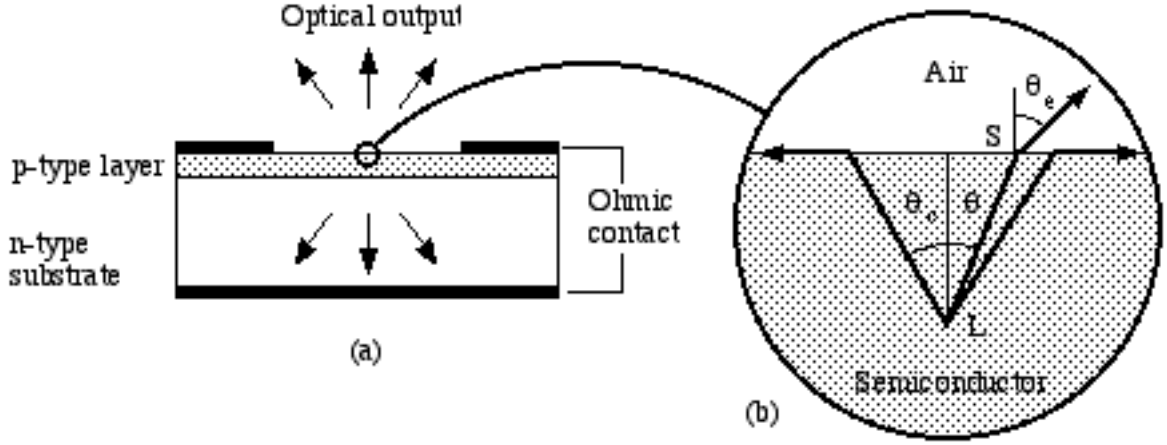


Figure 12.4-4 a) A typical surface emitting LED; b) geometry for calculation of external efficiency.

The proportion of the light that escapes into the air can be found as follows. We first define a source of radiation as the point L inside the LED and the point of escape from the surface of a given ray emitted by L as S, as shown in Figure 12.4-4b. If the angle made by this ray with the surface normal is θ, and the distance between L and S is r, the fraction of the radiation emitted by L lying in the solid angle between θ and θ + dθ is $2\pi r \sin(\theta)\, r\, d\theta / 4\pi r^2 = (1/2)\sin(\theta)\, d\theta$. The proportion of this fraction that actually crosses the interface is then given by $T(\theta) = 1 - \Gamma(\theta)^2$, where Γ(θ) is the Fresnel reflection coefficient for incidence at angle θ. The **external efficiency** η_e of the LED is then found as the integral of the product of the above two quantities, over the angular range for upward-travelling radiation, namely $0 \le \theta \le \pi/2$. We may therefore write:

$$\eta_e = {}_0\int^{\pi/2} \{T(\theta)/2\} \sin(\theta)\, d\theta \qquad 12.4\text{-}20$$

Now, T(θ) is a slowly-varying function, falling from a maximum value $T(0) = 4n/(n+1)^2$ at normal incidence (where n is the refractive index of the semiconductor) to $T(\theta_c) = 0$ at the critical angle. A reasonable approximation to Equation 12.4-20 is then:

$$\eta_e = T(0)/2\ {}_0\int^{\theta c} \sin(\theta)\, d\theta = T(0)/2\ \{1 - \cos(\theta_e)\} \qquad 12.4\text{-}21$$

Now, for typical LED materials, n is large (3.5 for GaAs), so $\theta_c = \sin^{-1}(1/n) \approx 1/n$ and $\{1 - \cos(\theta_c)\} \approx \theta_c^2/2 \approx 1/2n^2$. With these approximations, we may write the external efficiency as:

$$\eta_e \approx 1/n(n + 1)^2 \qquad 12.4\text{-}22$$

For GaAs, $\eta_e \approx 0.0141$, so that only about 1.4% of the light generated inside the LED can actually escape from the surface. Consequently, the DC external output is reduced to $P_e = P'\eta_e$. For the GaAs LED discussed earlier, the usable output would be 0.73 x 0.0141 mW ≈ 10 μW per mA drive current.

In fact, a fraction of the light is absorbed while travelling between L and S, reducing this figure still further. There are two partial solutions to the problem. Firstly, a reflector can be added to the lower surface, so that some of the light emitted downwards may be exploited. Secondly, the upper surface of the LED may be antireflection coated, which provides an improvement in T(0). Thirdly, the surface may be lensed (either by etching the surface into a hemisphere or by attaching an external microlens), so that a larger proportion of the light is incident on the interface at an angle less than θ_c. This solution is used extensively in display LEDs, but is inappropriate for lightwave systems, where the aim is to couple the LED output into an optical fibre. In this case, the power actually delivered into the fibre will be lower than the estimates obtained above by a factor η_c, the **coupling efficiency**.

Calculation of η_c for the general case of an optical fibre butt-coupled to the surface of an LED is a relatively complicated problem. However, for an emission area that is small compared with the fibre core, and arranged on-axis as shown in Figure 12.4-5, we can obtain a useful approximation to η_c.

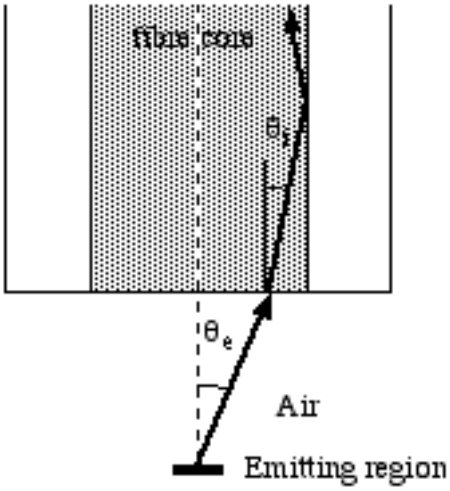


Figure 12.4-5 Geometry for calculation of LED: fibre coupling efficiency.

Light emitted by the LED travels in all directions, ranging from normal to the surface to parallel. It can be shown that the distribution of output intensity with external angle θe is Lambertian, following:

$$I(\theta_e) \approx I(0)\cos(\theta_e) \qquad 12.4\text{-}23$$

We may use this distribution perform a calculation very similar to the previous one. Firstly, we note that the fraction of the output lying in the solid angle between θ_e and $\theta_e + d\theta_e$ is:

$$dI = I(\theta_e)\sin(\theta_e)\,d\theta_e \,/\, {}_0\int^{\pi/2} I(\theta_e)\sin(\theta_e)\,d\theta_e$$

12.4-24

The proportion of this fraction that crosses the air/fibre interface is then given by $T'(\theta_e) = 1 - \Gamma'(\theta_e)^2$, where $\Gamma'(\theta_e)$ is the Fresnel reflection coefficient for incidence at an angle θ_e.

Now, of the part that is transmitted into the fibre, only the fraction that is total-internal-reflected at the core/cladding interface will actually be guided. In Chapter 8, we showed that this condition led to a maximum acceptance angle for a multimode step-index fibre of:

$$\theta_{emax} = \sin^{-1}\{\sqrt{(n_1^2 - n_2^2)}\} = \sin^{-1}(NA)$$

12.4-25

where NA is the numerical aperture of the fibre. Consequently, η_c may be found by multiplying Equation 12.4-23 by $T(\theta_e)$ and integrating from $\theta_e = 0$ to $\theta_e = \theta_{emax}$. Since θ_{emax} is typically small, and $T(\theta_e)$ is slowly-varying, η_c may be approximated by:

$$\eta_c = T'(0)\,{}_0\int^{\theta emax}\ I(\theta_e)\sin(\theta_e)\,d\theta_e \,/\, {}_0\int^{\pi/2} I(\theta_e)\sin(\theta_e)\,d\theta_e$$

12.4-26

Where $T'(0) = 4n'/(n'+1)^2$, and n' is the refractive index of the fibre core. At this point, we may substitute for $I(\theta_e)$ and perform the necessary integration. This yields the extremely simple result:

$$\eta_c = T'(0)\sin^2(\theta_{emax}) = T'(0)\,NA^2$$

12.4-27

Thus, for a step-index fibre with $n' \approx 1.5$ and $NA \approx 0.2$, we obtain a coupling efficiency of $\eta_c \approx 0.04$. This figure may be improved by contacting the fibre directly to the LED, or by using a lens between the two (if the emitting area is smaller than the fibre core). One alternative device (the edge emitting LED, which we will describe shortly) also offers an advantage. However, the fundamental cause of low external efficiency is the undirected nature of spontaneous emission. A major improvement may only be obtained by adopting the alternative mechanism of light production, stimulated emission.

DOUBLE HETEROSTRUCTURE LEDS

In place of the simple p-n junction, a more complicated structure is often used. Typically, this might be a P-n-N **double heterostructure**, containing three layers: a p-type layer with a wide energy gap, a narrow-gap n-type layer, and a wide-gap n-type layer. Now, we have already mentioned the way in which a single heterojunction may be used to provide potential barriers of different heights for electrons and holes; the double heterostructure may be used to provide a high barrier at different positions for electrons and holes. In this way, the recombination region may be limited to a defined region of space, resulting in an increase in the radiative recombination efficiency. We illustrate this in Figure 12.4-6, which shows the energy band diagram for a P-n-N structure.

Here the potential barrier faced by electrons is higher at the P-n interface, while that faced by holes is higher at the n-N interface. Under forward bias, electrons injected into the n-type layer from the right are prevented from diffusing away to the left by the first of these barriers. Similarly, holes injected into the n-type layer from the left are stopped from leaking away to the right by the second barrier. The net effect is that carriers may be poured into the central layer, which acts as a localised recombination region. Two further advantages stem from this construction. The first is that the larger band-gaps of the P- and N-type layers can make them effectively transparent to radiation generated in the narrow-gap recombination

region, thus reducing the amount of light that is absorbed before it escapes from the surface. The second is that the refractive index of the narrow-gap material at the centre of the structure may be higher than that of the material on either side. It is therefore possible to combine localisation of the recombination region with confinement of an optical field. This feature is exploited extensively in the double heterostructure laser, which will be described in the next Section.

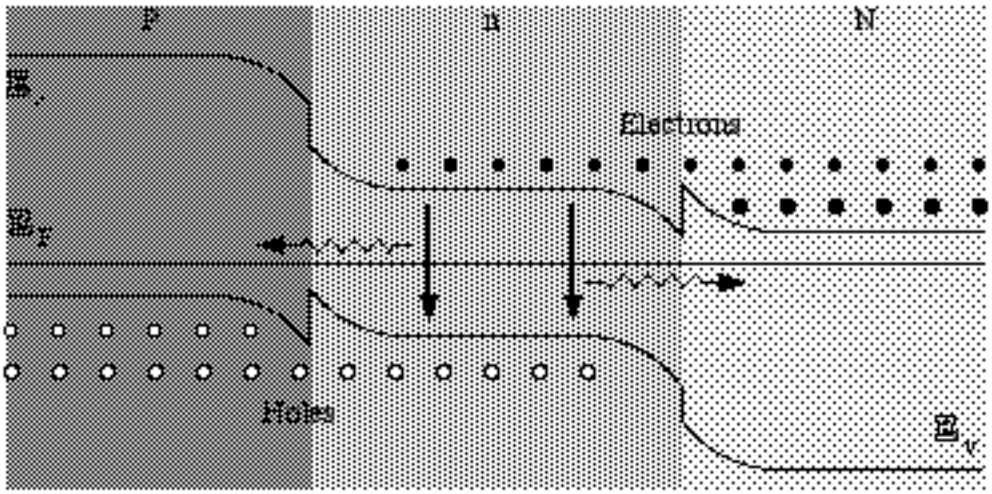


Figure 12.4-6 Energy band structure for a P-n-N double heterostructure in equilibrium.

Figure 12.4-7 shows a typical surface-emitting double-heterostructure LED, the **Burrus-type LED** (named after its inventor). This is a high-efficiency device, suitable for use with multimode fibre. To fabricate the device, a double heterostructure is first grown on a substrate, and a SiO_2 isolation layer is then deposited on the lower layer of the heterojunction. This layer is etched, to expose the hetero- junction over a small window. Metallization layers are then added to the upper and lower surfaces; clearly, the lower of these makes contact with the heterojunction only in the region of the window. A deep well, aligned with this window, is then etched right through the substrate to the upper layer of the heterojunction, and a multimode optical fibre is epoxied into the well in contact with the LED surface. When a forward current is passed through the LED, it flows mainly through the region immediately below the fibre. The recombination region (or **active region**) is therefore confined vertically by the double heterojunction and laterally by the distribution of current flow.

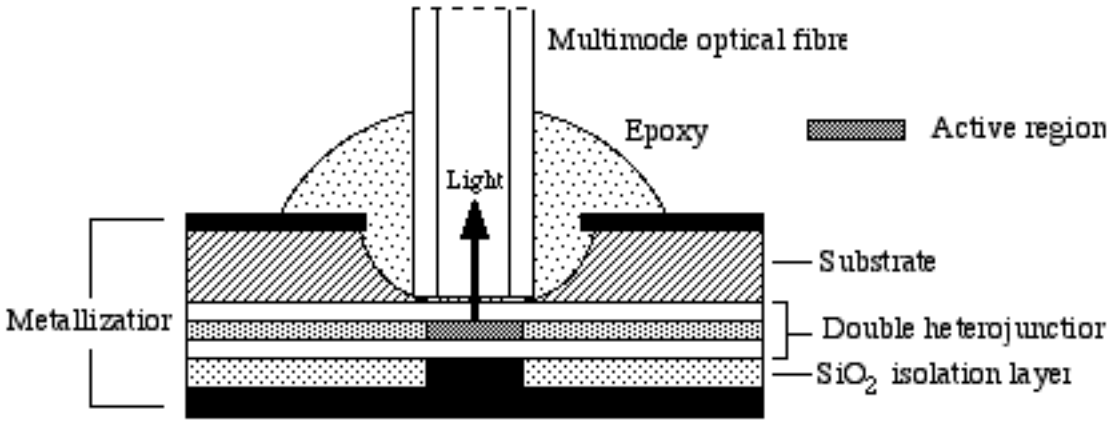


Figure 12-4-7 A Burrus-type double heterojunction surface-emitting LED.

THE EDGE-EMITTING LED

Figure 12.4-8 shows an alternative geometry, the **edge-emitting** or **superluminescent LED**. This is also based on a double heterostructure, but now the emission is taken from the edge of the junction rather than its surface. Generally, the current is forced to flow through a narrow strip down the device centreline by the introduction of a current-blocking silica layer. As a result, the active volume is constrained vertically by the double heterostructure and laterally by the current flow. Most importantly, light generated by spontaneous emission is also confined vertically to a certain extent; the refractive index difference at the top and bottom of the active layer results in total internal reflection. This tends to channel a large fraction of the light towards the emission windows at relatively shallow angles. While the radiation pattern is still Lambertian in the horizontal plane (with a half-power beam width of ≈ 120°), it is therefore considerably compressed in the vertical plane (into a width of ≈ 30°).

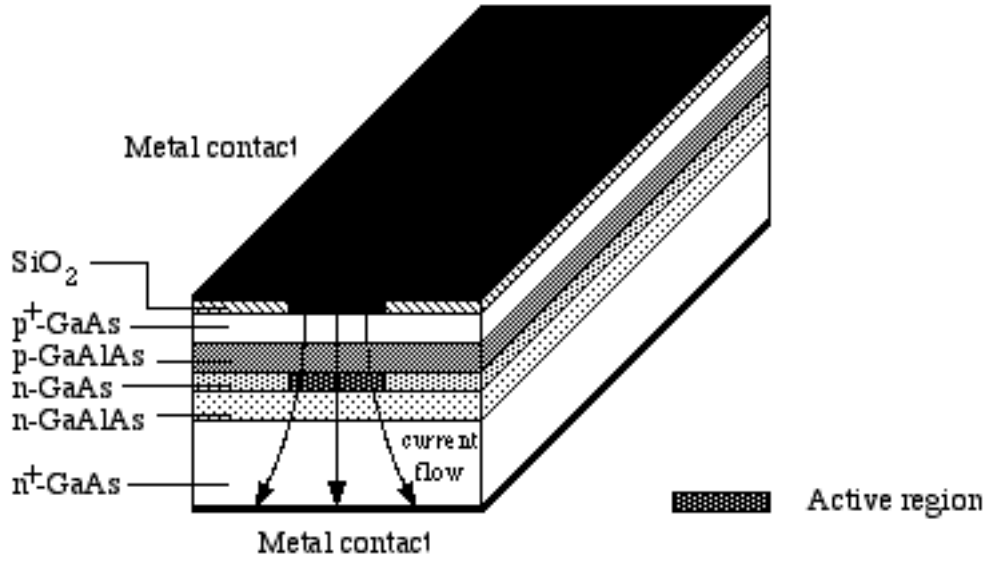


Figure 12.4-8 An edge-emitting LED.

The large distance travelled before the light escapes does result in an increased level of absorption (although this may be minimised by making the active layer thin, and forcing most of the light to travel in the wide-bandgap confining layers). Consequently, the external efficiency is lower than in a surface-emitting LED. However, this is more than compensated for by the increased directivity of the output, which allows much higher coupling efficiency to a fibre. In turn, this makes a reduction in the internal efficiency possible, thus improving the high-frequency response. Antireflection coatings may again be used to improve the transmission through the emitting facet. Edge-emitting LEDs may therefore have bandwidths of ≈ 5x their surface-emitting counterparts, while still coupling ≈ 5x the power into a low-NA fibre.

Despite all this, the performance improvement offered is only moderate, and a major difference can only be obtained using the alternative mechanism of light production, stimulated emission. Surprisingly, the structure required is very similar, but it has entirely different optical characteristics. We shall now consider this device - the **semiconductor laser** - in detail.

12.5 THE SEMICONDUCTOR LASER I: BASIC OPERATION

To explain the difference between the semiconductor laser and the LED, we must return to the results of Chapter 11. When absorption and stimulated emission are the dominant processes, the rate equation for a travelling photon wave in our two-level model may be taken as:

$$\partial\phi/\partial t = -v_g\partial\phi/\partial z + G\phi(n - n_0) \qquad 12.5\text{-}1$$

In the steady-state, $\partial\phi/\partial t = 0$, so that:

$$\partial\phi/\partial z = g_p\phi \qquad 12.5\text{-}2$$

where the power gain is $g_p = G(n - n_0)/v_g$. In this case, we can find the variation of ϕ with z as:

$$\phi(z) = \phi(0)\exp(g_p z) \qquad 12.5\text{-}3$$

Equation 12.5-3 describes an exponentially growing wave. This type of amplification ought to offer a method of light generation, if we can harness it. Clearly, we require $n > n_0$ for g_p to be positive, so that net gain occurs. To achieve this, we must find some means to raise the electron density far above the equilibrium level. The injection of electrons across a forward-biassed diode immediately suggests itself as a suitable method.

Figure 12.5-1 shows a three-dimensional sketch of the band-structure of a p-n^+ junction diode under heavy forward bias. Here the forward voltage V is so large that the relative levels of the conduction band on either side of the junction are the opposite of those in equilibrium. In this case, we would expect a veritable flood of electrons to be injected into the p-side. An optical wave travelling parallel to and sufficiently close to the depletion layer will then move through a region with greatly increased electron concentration.

Unfortunately, our previous expression for n_p' is inaccurate at the high forward bias required to invert the population. However, by a more accurate calculation, it can be shown that the net gain is obtained when the difference between the quasi Fermi-levels $\mathbb{E}_{Fn}$ and $\mathbb{E}_{Fp}$ is greater than the bandgap $\mathbb{E}_g$.

Now, the injected electrons will diffuse away from the depletion layer edge, so the gain must fall off in the same direction. Arguably, the excess electron concentration will fall to zero within about a diffusion length. Since no amplification will occur outside this region, the optical wave should be confined near the depletion layer to take full advantage of the gain. One approach would be to force it to propagate in a waveguide that passes through the gain region. We might then ask, how should the guide be designed for optimum gain?

To answer this, we must find the effective power gain due to a distribution $g_p(x)$ that varies over the guide cross-section. This can be done using the methods of Chapter 9, where we calculated the effect of a nonuniform index perturbation Δn on a guided mode in a phase modulator. There we found that the change in effective index was found from the overlap between the perturbation and the normalised transverse field, as $\Delta n_{eff} = < \Delta n , |E|^2 >$. It is easy to show that the effective power gain is also found as an overlap, this time between the transverse field and the gain distribution, i.e. as:

$$g_{peff} = < g_p , |E|^2 >$$

12.5-4

Consequently, we may deduce that the most important condition for high effective gain is a strong overlap between the modal field and the gain distribution.

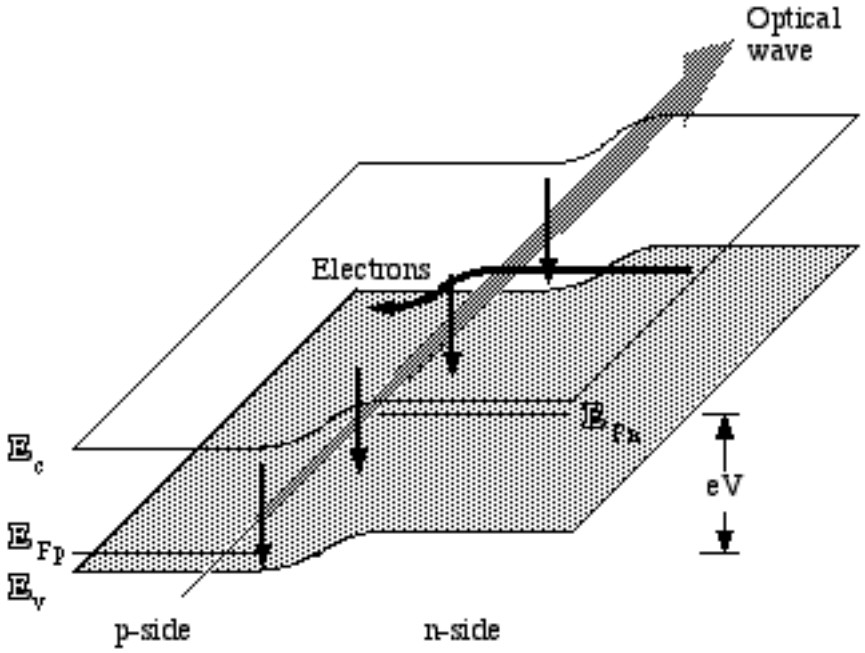


Figure 12.5-1 Amplification of a travelling optical wave near the depletion layer of a forward biased p-n junction.

Early semiconductor lasers were made from p-n homojunctions. These contain an in-built planar guide, formed by the increase in refractive index that occurs naturally in the depletion region due to the absence of free carriers. However, since the index change obtained this way is very small ($\Delta n/n \approx 0.1$ - 1%), such a guide provides a very weak confinement of the optical field. The lack of any kind of potential barrier also allows the diffusion of injected carriers away from the junction. Consequently, homostructure lasers have a low effective gain and are of historical interest only.

A much more successful geometry is the **double heterostructure laser**, which uses compositional variations to optimise the optical and carrier confinement. For example, Figure 12.5-2a shows the band diagram for a P-n-N heterostructure under strong forward bias. Note that there is a potential well in the n-region, in both the conduction band and the valence band. As mentioned earlier, electrons injected into the upper well from the N-side are highly localised, and available for recombination with holes injected into the lower well from the P-side.

Figure 12.5-2b shows the corresponding refractive index distribution across the junction; the wide band-gap P- and N-type regions have a lower index than the n-region, so a roughly symmetric slab guide (with $\Delta n/n \approx 5\%$) is created exactly in the high-gain region.

The modal field (Figure 12.5-2c) is then well confined, and has a strong overlap with the gain. Furthermore, the confinement is highly stable, since it is hardly altered by changes in the injection level. Modern semiconductor lasers are therefore almost exclusively built around variants of the double heterostructure. For example, in the **separate confinement heterostructure** (SCH), optical field confinement and localisation of the recombination

region are optimised separately, using a quadruple heterojunction. The two inner junctions limit the active region, while the two outer ones confine the optical field.

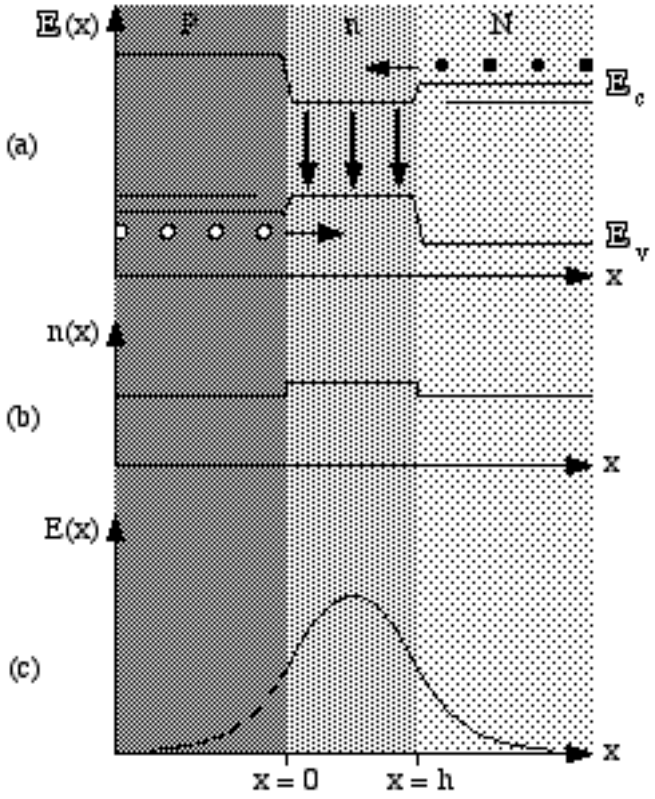


Figure 12.5-2 a) Energy bands, b) refractive index and c) modal field in a double heterostructure laser.

Figure 12.5-2b shows the corresponding refractive index distribution across the junction; the wide band-gap P- and N-type regions have a lower index than the n-region, so a roughly symmetric slab guide (with $\Delta n/n \approx 5\%$) is created exactly in the high-gain region. The modal field (Figure 12.5-2c) is then well confined, and has a strong overlap with the gain. Furthermore, the confinement is highly stable, since it is hardly altered by changes in the injection level. Modern semiconductor lasers are therefore almost exclusively built around variants of the double heterostructure. For example, in the **separate confinement heterostructure** (SCH), optical field confinement and localisation of the recombination region are optimised separately, using a quadruple heterojunction. The two inner junctions limit the active region, while the two outer ones confine the optical field.

DESIGN EXAMPLE

We may estimate the effective power gain in a double heterostructure laser as follows. In the central active layer, we may take the power gain to be constant, and equal to g_{p1}. However, there may also be additional sources of loss (free carrier absorption, for example), which we might define by a power absorption coefficient α_{p1}. In the unpumped outer layers, there will be no gain, only loss defined by α_{p2} and α_{p3} in the left- and right-hand layers, respectively. For the co-ordinates of Figure 12.5-2, the overlap integral in Equation 12.5-4 may then be evaluated as:

$$g_{peff} = \Gamma_1(g_{p1} - \alpha_{p1}) - \Gamma_2\alpha_{p2} - \Gamma_3\alpha_{p3} \qquad 12.5\text{-}5$$

where the overlap terms Γ_1, Γ_2 and Γ_3 are defined as:

$$\Gamma_1 = \int_0^h |E|^2 \, dx \quad , \quad \Gamma_2 = \int_{-\infty}^0 |E|^2 \, dx \quad , \quad \Gamma_3 = \int_h^\infty |E|^2 \, dx \qquad 12.5\text{-}6$$

and $\Gamma_1 + \Gamma_2 + \Gamma_3 = 1$. To maximise the effective gain, Γ_1 should be made much greater than Γ_2 and Γ_3, by confining as large a fraction of the optical field as possible inside the central pumped region. For a more complicated channel guide geometry, we might write by analogy:

$$g_{peff} = \Gamma(g_p - \alpha_{pcore}) - (1 - \Gamma)\alpha_{pclad} \qquad 12.5\text{-}7$$

Where Γ is the overlap of the guided mode with the active region - known as the **confinement factor**, and α_{pcore} and α_{pclad} are the losses in the core and cladding, respectively.

PRACTICAL LASER GEOMETRIES

An optical amplifier may be turned into an oscillator by providing positive feedback. We have already described how this is done in Chapter 4; all that is needed is to add mirrors to the gain block, so that the structure becomes a Fabry-Perot cavity. In the semiconductor laser, the mirrors may be fabricated very simply, by cleavage down crystal planes orthogonal to the junction plane, as shown in Figure 12.5-3. Naturally, their reflectivity is fixed; for a material of refractive index 3.5, surrounded by air, the amplitude reflectivity is $R_1 = R_2 \approx (3.5 - 1) / (3.5 + 1) \approx 0.56$. This figure is high enough to obtain a suitable level of feedback, while still allowing a good fraction of the light generated to escape from the cavity. Gain is provided by the injection of current in a direction normal to the junction. Because a planar structure provides confinement of the light only in one direction, we would expect light to travel between the mirrors as a sheet beam, growing in amplitude as in propagates.

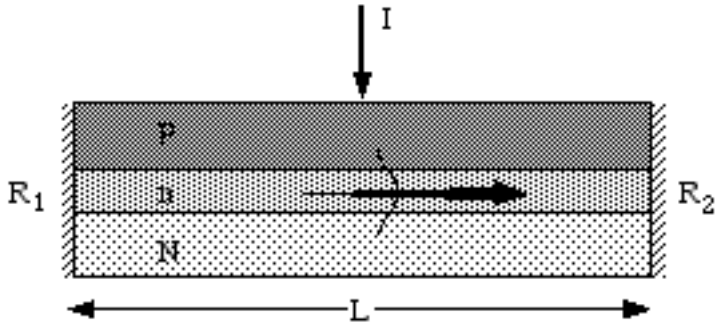


Figure 12.5-3 Basic geometry of a planar double heterostructure laser.

In practise, the use of a simple layered structure is undesirable. Firstly, both the drive current and the heat generated by non-radiative transitions will be large, since they are proportional to the area of the junction. Secondly, the finite lateral extent of the guide implies that it will actually function as a very broad, multimoded channel guide. The optical field inside the cavity will then consist of a random mixture of modes, which may lead to a highly nonuniform intensity distribution across the beam. Thirdly, broad-area emission will result in very low coupling efficiency into a fibre. Almost always, extra refinements are used to reduce the active volume and introduce transverse confinement, so that the light propagates in a single-moded channel guide.

The primary problem with semiconductor lasers is therefore to combine two-dimensional confinement of the optical wave with localisation of the recombination region. There are many possible solutions. One is to restrict the flow of current laterally, for example by using an SiO_2 isolation layer with a stripe contact window, as shown earlier in Figure 12.4-8. This geometry is known as the **oxide stripe laser**. Alternatively, a full-area contact might be used, but the semiconductor might be made highly resistive in all except a narrow stripe region, for example by proton bombardment. Both methods restrict the gain region laterally. However, the lateral optical confinement is only moderate, since this is introduced by a local increase in gain rather than by an index variation. Such lasers are therefore often described as **gain-guided**. Due to the weak confinement, the guide may support several transverse modes, and the relative importance of each mode may be highly dependent on the level of injection. For stable operation on a single transverse mode, stronger lateral confinement must be established using refractive index changes. This requires a more complicated structure.

Figure 12.5-4 shows the simplest example of a suitable geometry, the **ridge waveguide laser**. In this example, the laser is designed to emit long-wavelength IR radiation (at $\lambda_0 \approx 1.5$ μm), so the InP/InGaAsP materials system has been used. The starting point is a planar InP-InGaAsP-InP double heterostructure, but the upper InP layer is partially etched into a rib (typically ≈ 5 μm wide) to provide lateral confinement of the mode by strip-loading. Note that the current is also automatically confined by the finite width of the rib, although it may start to spread as it passes into the active layer.

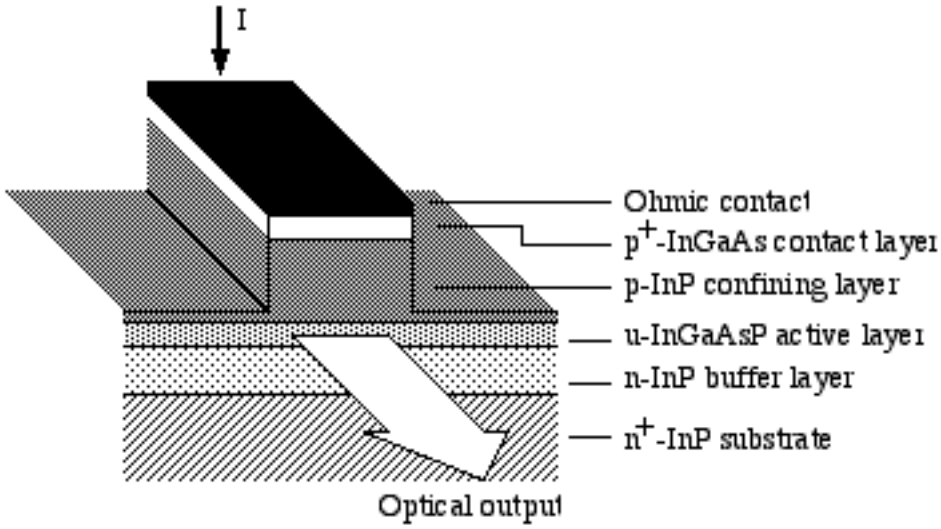


Figure 12.5-4 A ridge waveguide laser (after A.W. Nelson et al., Br. Telecom J. 4, 85-103 (1986)).

Figure 12.5-5 shows a more complicated example, the **buried heterostructure laser**. The laser shown is designed to emit at a shorter wavelength ($\lambda_0 \approx 0.83$ μm), and is therefore based on the GaAs/GaAlAs materials system. The starting point is a GaAlAs-GaAs-GaAlAs double heterostructure, which is again fashioned into a ridge. This time, however, the ridge is etched completely down to the substrate, and buried on either side by additional high-resistivity GaAlAs. Lateral optical confinement is not lost, however, since the burying layer has a lower refractive index than the active region. Contact is then made to the top of the heterostructure via an oxide stripe window, and the current is constrained to flow through the active region by the high resistivity of the surrounding material.

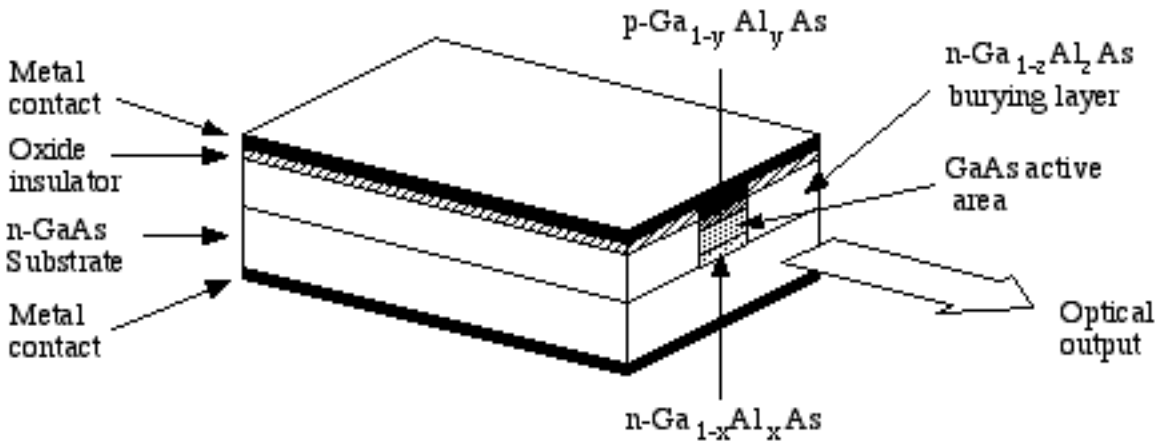


Figure 12.5-5 A GaAs/GaAlAs buried heterostructure laser

Still better current confinement is possible using more complicated structures. For example, Figure 12.5-6 shows the **double-channel planar buried heterostructure laser**. This begins life as a planar InP-InGaAsP-InP double heterostructure, which has two deep channels etched through it to the substrate. Four additional layers are then added, by further epitaxial growth. In order, these consist of p-InP, n-InP, p-InP and p^+-InGaAs. The important feature is that by utilising a peculiarity of liquid- phase epitaxial growth known as **selective area epitaxy**, the n-InP layer may be constrained to grow everywhere except over the central ridge. As a result, the structure contains several in-built, reverse-biased p-n junctions, which provide far better current blocking than high-resistivity material. Not only do they force the current to flow into the top of the ridge, but they also prevent it from spreading out of the sides of the active region.

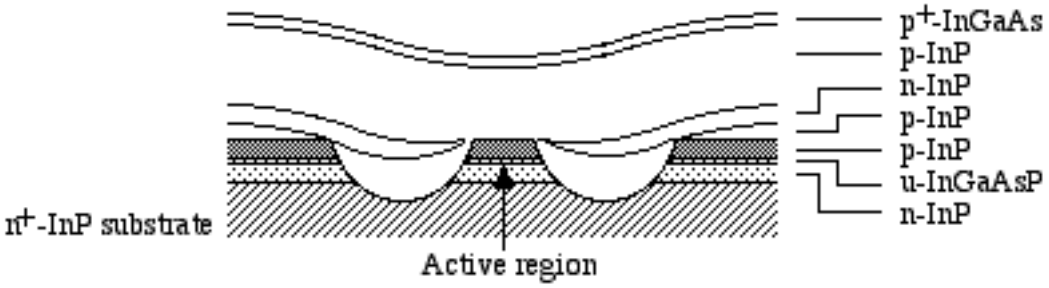


Figure 12.5-6 Cross section of an InP/InGaAsP double-channel planar buried heterostructure laser (after A.W.Nelson et al., Br. Telecom J. 4, 85-103 (1986)).

CONDITIONS FOR LASER OPERATION

We now consider the basic analysis of semiconductor lasers. From Chapter 4, we know that two conditions are required for laser operation. The first is that a longitudinal resonance is satisfied, so that the phase-change accumulated in a round-trip up and down the cavity is a whole number of multiples of 2π. Only then will the feedback be positive, so that the amplitude of the optical field inside the cavity is high. For a single-mode guide with effective index n_{eff}, this condition may be written as:

$$2k_0 n_{eff} L = 2\nu\pi \qquad 12.5\text{-}8$$

However, it should be noted that the effective index is normally dependent on the level of injection, due to the free carrier contribution to the dielectric constant. The resonant wavelengths are spaced by:

$$\Delta\lambda = \lambda_\nu - \lambda_{\nu+1} = \lambda_\nu^2/2Ln_{eff} \qquad 12.5\text{-}9$$

$\Delta\lambda$ is therefore inversely proportional to the cavity length. The second condition for laser operation is that the round-trip gain must exceed unity, so that:

$$R_1R_2 \exp(2g_{eff}L) > 1 \qquad 12.5\text{-}10$$

The effective power gain at the threshold of lasing is therefore:

$$g_{pefft} = 2g_{eff} = 1/L \log_e(1/R_1R_2) \qquad 12.5\text{-}11$$

Consequently, the shorter the cavity, the higher the threshold power gain. For a cavity of length 250 μm, with end-mirrors of reflectivity $R_1 = R_2 = 0.56$, we obtain $g_{pefft} \approx 4{,}640\ m^{-1}$. Note that the gain that must be provided in the active region is even higher, since from Equation 12.5-7 we obtain:

$$g_{pt} = \alpha_{pcore} + \{(1 - \Gamma) / \Gamma\}\alpha_{pclad} + (1/\Gamma L) \log_e(1/R_1R_2) \qquad 12.5\text{-}12$$

In our earlier discussion of LEDs, we described how the distribution of electron and hole energies leads to a finite spectrum for spontaneous emission. For the same reason, gain can be obtained in a semiconductor laser over a finite range known as the **gain bandwidth**. Normally, the variation of gain with wavelength is written in the form:

$$g_p(\lambda) = a(n - n_0) - b(\lambda - \lambda_p)^2 \qquad 12.5\text{-}13$$

where a and b are constants, and λ_p is the wavelength for peak gain. With this in mind, we would expect the output of a laser with a relatively long cavity to consist of a series of closely-spaced lines, as shown in Figure 12.5-7a. Each corresponds to a particular longitudinal mode, when Equation 12.5-8 is satisfied for a given value of ν. The gain bandwidth of (for example) an InGaAsP laser has a half-width of ≈ 30 nm. For the typical parameters of $\lambda_\nu = 1.5$ μm, L = 250 μm and $n_{eff} = 3.2$ we obtain $\Delta\lambda \approx 1.4$ nm, so many lines must lie within the gain bandwidth. We cannot comment on the relative powers in each of the lines yet (except to say that there will be more power in those nearest to the gain peak) since the details of the spectrum depend on the level of injection. However, an extended spectrum is generally undesirable in fibre communications systems.

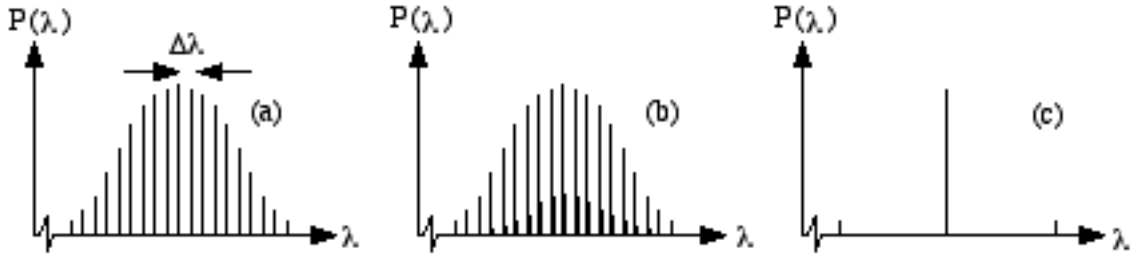


Figure 12.5-7 Representative output spectra for a) and b) a long-cavity laser, with one and two transverse modes, respectively; c) a short-cavity laser supporting a single transverse mode.

Furthermore, we note that a laser with a broader stripe that supports several transverse modes would have a more complicated spectrum. Since each transverse mode will have a different effective index, each will have its own associated spectrum of longitudinal resonances. However, the effective gain of higher-order modes is likely to be lower, since

they are more weakly confined than the fundamental mode. Consequently, the spectrum of a two-moded laser might be as shown in Figure 12.5-7b, where the smaller lines correspond to the second-order mode. Since the lowest-order mode is generally the most useful, single-transverse-mode operation is highly desirable. In this case, and with a much shorter cavity, we might expect only one longitudinal resonance to fall inside the gain bandwidth, so that the output is effectively single-line (Figure 12.5-7c). However, this requires an extremely small value of L ($\approx 50\ \mu m$).

QUANTUM WELL AND MULTIQUANTUM WELL LASERS

Band-gap engineering offers further scope for the control of laser properties. For example, Figure 12.5-8a shows the equilibrium band diagram of an undoped GaAlAs-GaAs-GaAlAs double heterostructure. Due to the difference in bandgap between the materials, there are potential wells in both the conduction band and the valence band. In conventional lasers, these are exploited purely for carrier confinement. However, as the width h of the central layer is reduced, new effects start to occur. In Chapter 11, we analysed the confinement of an otherwise free electron in a one-dimensional well, and showed that the result was a quantization of the electron energy. For an infinite well of width h, with a minimum potential energy of $\mathbb{V} = 0$, the allowed energies were:

$$\mathbb{E} = \nu^2(\pi^2 hbar^2/2mh^2) + hbar^2\beta^2/2m \quad (\text{where } \nu = 1, 2 \ldots) \qquad 12.5\text{-}14$$

Of these two contributions, the first arises from the quantization effect of the well, while the second is a kinetic energy term due to motion in the two remaining allowed directions. By analogy, we might expect the energies of conduction electrons in the well of Figure 12.5-8 to be of the form:

$$\mathbb{E}_{c1}' = \mathbb{E}_{c1} + \nu^2(\pi^2 hbar^2/2m_e^*h^2) + hbar^2\beta^2/2m_e^* \quad (\text{where } \nu = 1, 2 \ldots) \qquad 12.5\text{-}15$$

while the corresponding hole energies might be:

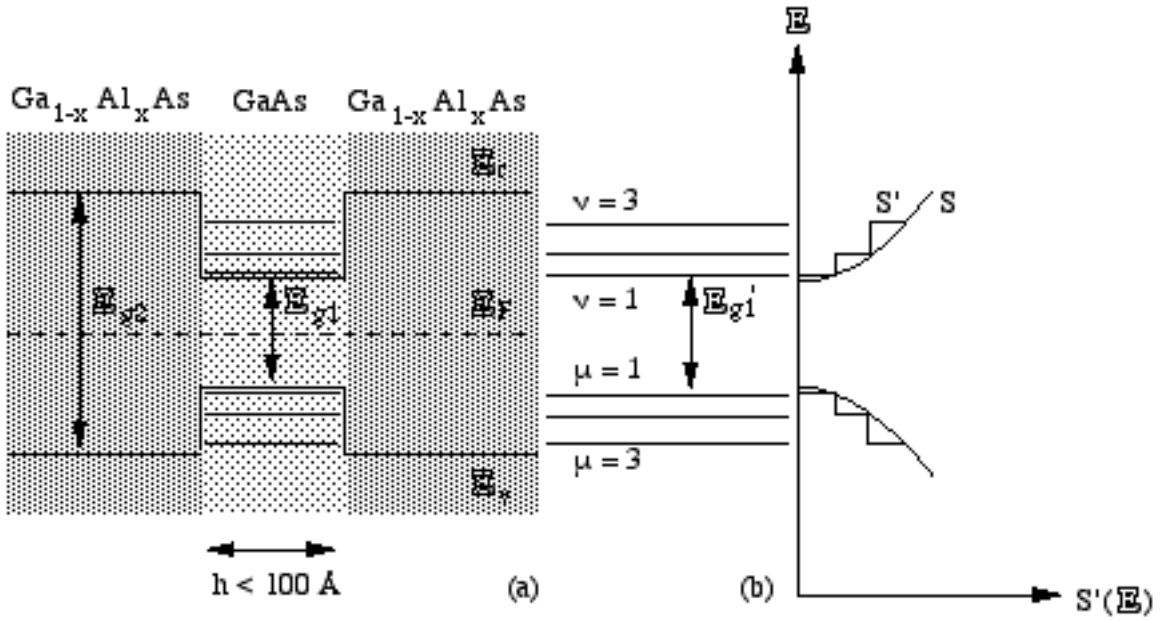
$$\mathbb{E}_{v1}' = \mathbb{E}_{v1} - \mu^2(\pi^2 hbar^2/2m_h^*h^2) - hbar^2\beta^2/2m_h^* \quad (\text{where } \mu = 1, 2 \ldots) \qquad 12.5\text{-}16$$

where $\mathbb{E}_{c1}$ and $\mathbb{E}_{v1}$ are the extremities of the conduction and valence bands in the GaAs layer, and m_e^* and m_h^* are the electron and hole effective masses. As a result, the minimum difference between conduction and valence levels is no longer $\mathbb{E}_{g1} = \mathbb{E}_{c1} - \mathbb{E}_{v1}$, but:

$$\mathbb{E}_{g1}' = \mathbb{E}_{g1} + (\pi^2 hbar^2/2h^2)\ (1/m_e^* + 1/m_h^*) \qquad 12.5\text{-}17$$

This energy difference corresponds to a transition between stationary electrons and holes in the levels defined by $\nu = 1$, $\mu = 1$. This is equivalent to the creation of an artificial semiconductor, with a bandgap different from that of either of the materials forming the well. Note that $\mathbb{E}_{g1}'$ is controlled essentially by the value of h. Consequently, the emission wavelength of a laser constructed in this way should be similarly variable.

In fact, the wavelength may be tuned over almost the entire range set by the bandgaps of the two constituent materials, GaAs and $Ga_{1-x}Al_xAs$. The tuning range is therefore limited mainly by the transition of GaAlAs to an indirect bandgap material at $x \approx 0.35$. Using quantum wells made from InP and InGaAs (which are both direct-gap), the wavelength may be tuned over the staggering range 1.07 - 1.55 μm, by varying the well width between 10 and 110 Å.



12.5-8 a) Band structure and b) density of states for a GaAlAs-GaAs-GaAlAs quantum well.

The quantum size effect offers other advantages besides wavelength tunability. For example, it is possible to repeat the calculation to find the density of states S'(𝔼) in the conduction band inside a single well. If this is done, we obtain:

$$S'(\mathbb{E}) = \nu m_e^*/\pi h\hbar^2 \quad \text{for} \quad \mathbb{E}_\nu \le \mathbb{E} \le \mathbb{E}_{\nu+1} \qquad 12.5\text{-}18$$

where $\mathbb{E}_\nu = \mathbb{E}_{c1} + \nu^2(\pi^2\hbar^2/2m_e^*h^2)$. Consequently, S'(𝔼) is discontinuous, as shown in Figure 12.5-8b. Although this is very different from our previous parabolic result for S(𝔼) - which is also the density of states for an infinite well width - the most important point is that S'(𝔼) intersects S(𝔼) at each energy $\mathbb{E}_\nu$, but otherwise lies inside it. We may therefore deduce that there must be a smaller total density of states in the conduction band in a quantum well (actually, this is obvious, since the effect of quantization is to disbar many possible states). Now, remember that we must invert the electron population to obtain optical gain. With fewer states to invert, the electron concentration required for transparency must be lower, so a quantum well laser must have a lower threshold than a conventional one. However, its peak gain must be reduced, since the conduction band will be filled more easily.

If the well is large enough, a double heterostructure laser can be made using a single quantum well as the active layer. However, as the well size is reduced, it becomes impossible to provide suitable optical confinement - the resulting waveguide is too narrow, so that the optical field extends considerably outside the gain region. It is then normal to construct the active layer as a multilayer stack, consisting (for example) of many alternating layers of InP and InGaAs, and to provide the optical confinement separately. In this case, the structure is known as a **multi-quantum well** (MQW) laser.

12.6 THE SEMICONDUCTOR LASER II: STEADY-STATE ANALYSIS

Assuming that the laser operates on a single transverse mode, and that the output is concentrated in a single line, we may model the operation of a laser using relatively minor

modifications of the two-state, lumped-element rate equations derived earlier for LEDs. Although the optical field inside the cavity will actually consist of a standing-wave pattern, we shall ignore this and simply assume an <u>average</u> value for the photon density ϕ. This will be high at the resonant wavelength (which we take to be λ_g), and zero otherwise. For simplicity, we shall ignore optical loss, and assume a confinement factor of $\Gamma = 1$. We may then take the photon rate equation for a cavity of volume v in the form:

$$d\phi/dt = \beta n/\tau_{rr} + G\phi(n - n_0) - \phi/\tau_p \qquad 12.6\text{-}1$$

Here we have included terms describing spontaneous emission ($\beta n/\tau_{rr}$), stimulated emission and absorption ($G\phi(n - n_0)$), and loss of photons from the cavity (ϕ/τ_p). The factor β has been introduced into the first of these to describe the fraction of the isotropically generated spontaneous emission which is actually coupled into the guided mode; typically, this is very small ($\approx 10^{-3}$), due to the low NA of a single-mode guide. Similarly, the electron rate equation may be taken as:

$$dn/dt = I/ev - n/\tau_e - G\phi(n - n_0) \qquad 12.6\text{-}2$$

Since Equations 12.6-1 and 12.6-2 are non-linear, we will solve them by means of approximations that are valid in the two important regimes of operation. In the steady state, the left-hand side of each of these equations is zero, so we may then combine them to eliminate $G\phi(n - n_0)$. This yields:

$$\phi/\tau_p = \beta n/\tau_{rr} + (I/ev - n/\tau_e) \qquad 12.6\text{-}3$$

Now, the light flux out of the cavity is $\Phi = v\phi/\tau_p$ as before. As we have ignored optical loss, this flux must all emerge from the cavity end-mirrors. If each photon carries energy hc/λ_g, the optical power output is then $P = hc\Phi/\lambda_g$, or:

$$P = (hcv/\lambda_g)\{\beta n/\tau_{rr} + (I/ev - n/\tau_e)\} \qquad 12.6\text{-}4$$

Note that Equation 12.6-4 represents the <u>total</u> output, so that P/2 emerges from each of the mirrors.

OPERATION BELOW THRESHOLD

For a low photon density, spontaneous emission is the dominant process, so we may set the term $\{I/ev - n/\tau_e\}$ (which describes stimulated emission) equal to zero. In this case, we have $n = I\tau_e/ev$, so that:

$$P = \eta hcI/e\lambda_g \qquad 12.6\text{-}5$$

where $\eta = \beta\tau_e/\tau_{rr}$. For low currents, the output is therefore proportional to current, as in the left-hand part of Figure 12.6-1a. The laser acts as an LED in this regime, with a reduced internal efficiency.

THE THRESHOLD CONDITION

Now, the existence of spontaneously emitted light in the laser cavity provides a source of optical noise, from which laser oscillation may start. The noise will travel up and down the cavity, and, if there is net gain, it will be amplified as it propagates. Very quickly, the photon density will become high, so that stimulated emission becomes the dominant process. In the

steady state, assuming that spontaneously emitted light is now negligible, we may approximate Equation 12.6-1 by:

$$G\phi(n - n_0) - \phi/\tau_p = 0 \qquad 12.6\text{-}6$$

We may then obtain the electron density during lasing as:

$$n = n_0 + 1/G\tau_p \qquad 12.6\text{-}7$$

As a result, the electron concentration is entirely independent of ϕ. Equation 12.6-7 must also represent the electron density at threshold, nt, and therefore describes the **threshold condition**.

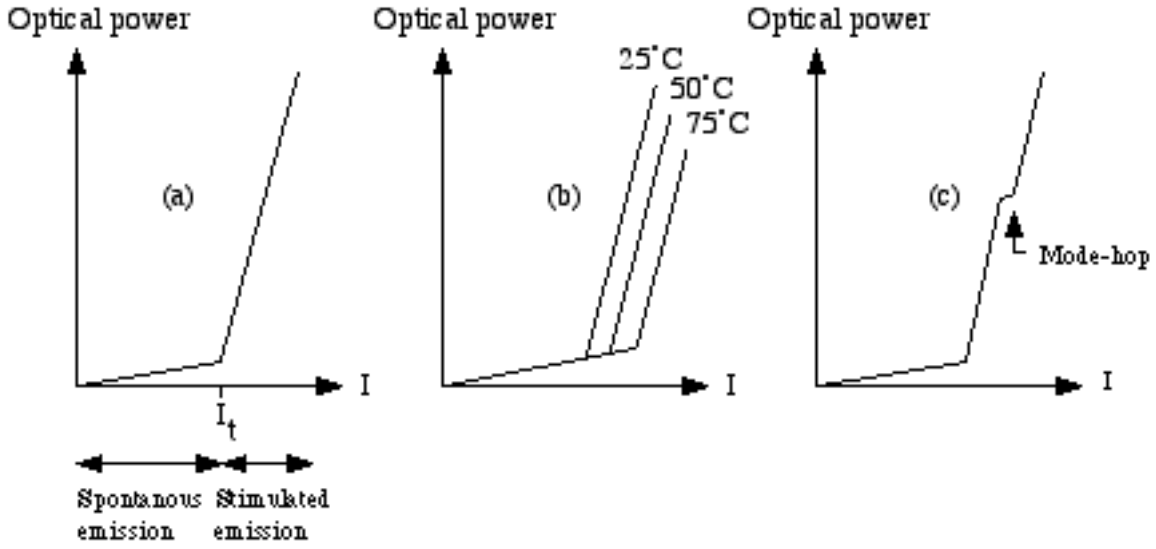


Figure 12.6-1 a) Optical power - current characteristic of a semiconductor laser, b) dependence of threshold current on temperature, and c) kink in characteristic due to mode-hopping.

DESIGN EXAMPLE

We may the result above to estimate the photon lifetime. Since we have already established the threshold condition, we may find n_t by an alternative route. Setting $\Gamma \approx 1$ and $\alpha_{core} = \alpha_{clad} \approx 0$ in Equation 12.5-12, the threshold gain is $g_{pt} \approx 1/L \log(1/R_1R_2)$. Since $g_{pt} = G(n_t - n_0)/v_g$, we may find the threshold electron density as:

$$n_t = n_0 + (v_g/GL) \log_e(1/R_1R_2) \qquad 12.6\text{-}8$$

Comparing Equations 12.6-7 and 12.6-8 we may then obtain the photon lifetime τp as:

$$\tau_p = L / \{v_g \log_e(1/R_1R_2)\} \qquad 12.6\text{-}9$$

The photon lifetime is therefore independent of the level of injection, and depends only on the cavity length and mirror reflectivity. For a semiconductor of refractive index n, we may put $v_g \approx c/n$. Assuming that n = 3.5, the end-mirror reflectivity is $R_1 = R_2 \approx 0.56$. If the cavity length is 250 μm, the photon lifetime is $\tau_p = 250 \times 10^{-6} \times 3.5 / \{3 \times 10^8 \times \log_e(1/0.562)\} = 2.5 \times 10^{-12}$ sec, or 2.5 psec.

OPERATION ABOVE THRESHOLD

We now consider operation above threshold. Assuming negligible spontaneous emission, Equation 12.6-4 approximates to:

$$P = (hcv/\lambda_g)\,(I/ev - n/\tau_e) \qquad 12.6\text{-}10$$

Since we have already shown that the electron concentration is clamped above threshold, n/τ_e must be constant. We might define it to be equal to I_t/ev, where I_t is a new parameter known as the **threshold current**. It is therefore proportional to the active volume v and the threshold electron density n_t, and inversely proportional to τ_e. Its exact value is highly dependent on the laser type; typical values are the range 10 - 100 mA. Using this definition, Equation 12.6-10 may be written as:

$$P = hc\{I - I_t\}/e\lambda_g \qquad 12.6\text{-}11$$

The output of a semiconductor laser is therefore a discontinuous function of the drive current, being proportional to $I - I_t$ in the lasing regime as shown in the right-hand part of Figure 12.6-1a. The peak output power is normally limited either by mirror damage (caused by the high optical intensity passing through the end-facets) or by thermal breakdown (caused by the high current density passing through the junction). Ouput powers in the range 1 - 10 mW are routinely available at all the major wavelengths (near 0.85 μm for GaAs/GaAlAs lasers, and 1.3 and 1.5 μm for InP/InGaAsP).

Generally, I_t rises with temperature, because the spread of energies in the conduction and valence band increases, reducing the density of carriers that can generate a photon of a given wavelength. The recombination time τe is also decreased. Consequently, the power - current characteristic of a laser at three values of T might be as shown in Figure 12.6-1b. In fact, operation at a greatly reduced temperature was often used to lower the threshold current to an acceptable level in early lasers, which had poor confinement factors. τ_e also decreases as the active region ages, so I_t rises with time.

MULTILINE OPERATION

The two-state model we have used cannot describe multiline operation. However, we may obtain a feel for the likely behaviour, as follows. If the spontaneous emission term is retained in Equation 12.6-1, a steady-state solution for the photon density may be found as:

$$\phi = (\beta n/\tau_{rr}) \,/\, \{1/\tau_p - G(n - n_0)\} \qquad 12.6\text{-}12$$

This shows that the light in the cavity may be considered to arise from spontaneously-emitted noise, 'amplified' by the factor $F = 1 \,/\, \{1/\tau_p - G(n - n_0)\}$. Since the denominator of this expression will be very small, F must be very large. In a more complicated model, including several longitudinal modes, we might assign a different gain to each, following Equation 12.5-13. We might then write:

$$F_\nu = 1 \,/\, \{1/\tau_p - [G(n - n_0) + b'(\lambda_\nu - \lambda_p)^2]\} \qquad 12.6\text{-}13$$

for the amplification factor for mode ν, where λ_ν is the wavelength of the mode and $b' = bv_g$. The relative sizes of each F_ν then depends on the closeness of each term $G(n - n_0) + b'(\lambda_\nu - \lambda_p)^2$ to $1/\tau_p$. For low currents (i.e., when $n - n_0$ is small), all the amplification factors are small, and roughly equal, so that multiline operation occurs. As $n - n_0$ increases, every value

of F_v will increase, but the line whose wavelength is closest to λ_p will have a proportionately larger amplication factor.

We can then guess that the mode spectrum will tend to alter as shown in Figure 12.6-2, where the power is gradually concentrated into one dominant line as the current rises. This is highly desirable, since it reduces the effect of dispersion when a laser is used as a source for optical communications. Note that an increase in drive current may also make it possible for the transverse lasing mode to change, for example, from the lowest-order mode at low current to the second-lowest at higher current. This behaviour is known as **mode-hopping**. It may be observed as a kink on the laser output characteristic, as shown in Figure 12.6-1c, and is generally undesirable.

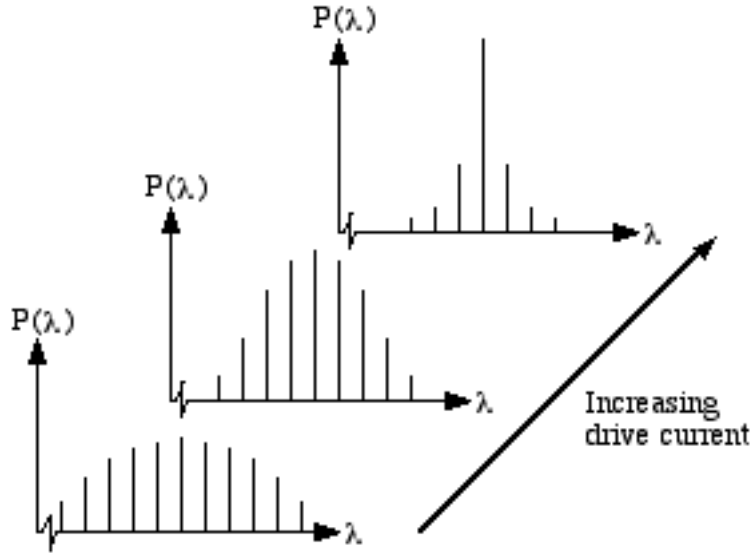


Figure 12.6-2 Typical variation of the mode spectrum of a semiconductor laser.

OPTICAL CHARACTERISTICS

We now consider briefly a number of other characteristics of a semiconductor laser. Firstly, in order to obtain good overlap between the optical field and the gain distribution, the height h of the active region of a double heterostructure laser supporting a single transverse mode is normally very small, of the order of 0.1 - 0.5 μm, while the corresponding width w is typically ≈ 1 - 5 μm. The core of the laser waveguide is therefore rectangular, with an aspect ratio of ≈ 10 : 1. The mode supported by the guide is similarly distorted, being highly compressed perpendicular to the plane of the layering, as shown in the near-field pattern of Figure 12.6-3.

To a good approximation, each facet behaves as a point emitter, so that the far-field radiation pattern is roughly that of a paraxial spherical wave. Now, we saw in Chapter 4 that the spread of a bounded beam is inversely dependent on its width. Due to the shape of the near-field pattern, we would expect the angular spread to be greatest in the plane perpendicular to the junction (also as shown in Figure 12.6-3). A typical value for the full-angle, half-maximum spread in intensity in this plane is $\theta_\perp \approx 30 - 50^o$; the corresponding spread $\theta_{||}$ parallel to the junction is much smaller, in the range 10 - 15°, but both figures depend strongly on the construction of the laser. The output may be collimated using a spherical lens, simply by placing the emitting area at the focal point. The resulting beam will roughly parallel, but (due to the differences in beam divergence) its cross-section will be elliptical. Approximate circular symmetry may be restored by passing the beam through an

anamorphic expansion telescope that magnifies in one direction only. However, the presence of gain in the active region causes a slight astigmatism of the modal phase-front. If accurate collimation is required, this must be corrected by passing the beam through an additional cylindrical lens.

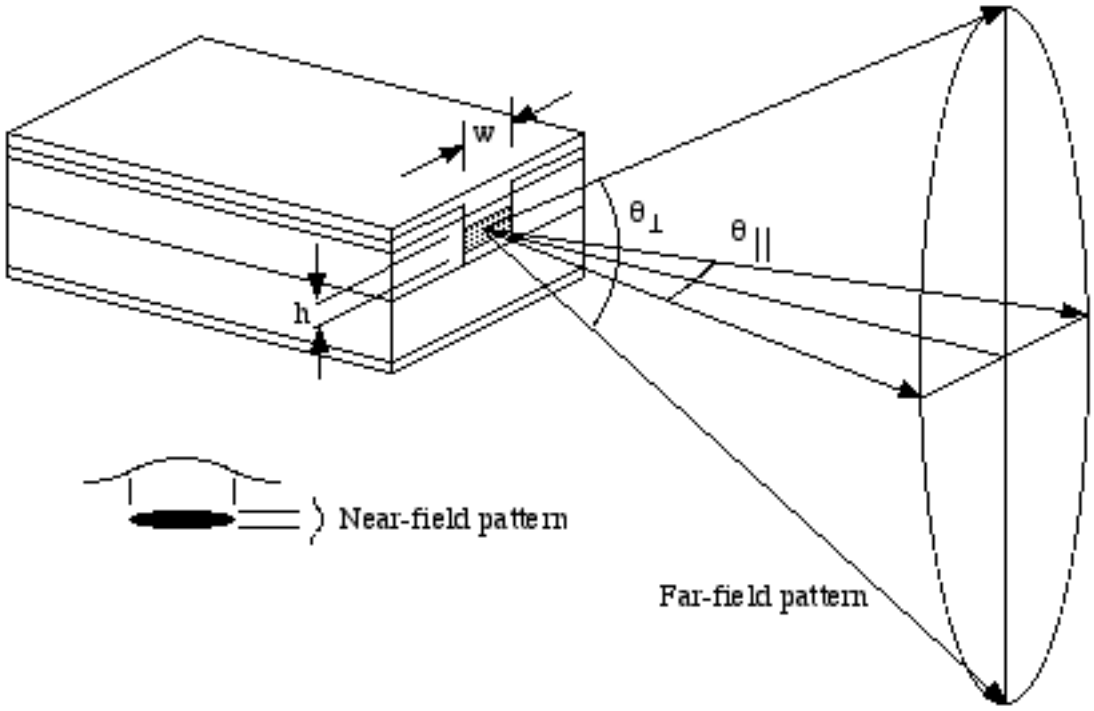


Figure 12.6-3 Near- and far-field radiation pattern from a buried heterostructure laser.

Because of the increased directivity and coherence of the light generated by stimulated emission, much higher coupling efficiency to a low-NA single-mode fibre can be obtained with a laser than with an LED. The coupling efficiency may be found by evaluating the overlap between the modal fields of the laser and fibre, much as was done in Chapter 8. Since the laser mode is typically rather small, an improvement in efficiency may be obtained by using a lens. This magnifies the laser mode slightly, so that it makes a better match to the fibre mode.

Finally, we note that a nominally single-moded guide may support two orthogonal polarization modes. However, due to differences in end-mirror reflectivity, the mode polarized parallel to the junction plane experiences greater round-trip gain, and will lase preferentially. Consequently, the stimulated emission is linearly polarized in this direction. However, spontaneous emission will add unpolarized light to the output, which will also have a component perpendicular to the junction plane. This lowers the degree of polarization, i.e. the ratio of orthogonally polarized components. This is limited, and will vary with the output power. A typical value for a laser driven well past threshold is greater than 50 : 1.

12.7 THE SEMICONDUCTOR LASER III: MODULATION

We now consider the behaviour of a semiconductor laser when it is modulated by a sinusoidal time-varying drive current, to demonstrate the improvement in its high-frequency response over that of an LED. Now, it can be shown that the time taken to build up the photon density in the cavity from zero is relatively long, due to the slow response of the

device at low current (when it operates as an LED). For high-speed operation, a laser is therefore normally biased past threshold, using a DC bias current. We shall therefore assume that the drive current is of the form $I = I' + I'' \exp(j\omega t)$, where I' is the DC bias, and I'' and ω are the amplitude and angular frequency of the superimposed AC modulation, respectively. Note that the bias is chosen so that $I' > I_t$ and $I' \gg I''$. We shall also define the other quantities of interest in a similar way, as the sum of a DC bias and an AC modulation term. We therefore put $\phi = \phi' + \phi''\exp(j\omega t)$, $n = n' + n'' \exp(j\omega t)$ and $P = P' + P''\exp(j\omega t)$.

We now substitute these expressions into Equations 12.6-1 and 12.6-2. To simplify the results, we perform the following operations: (1) we neglect spontaneous emission as small by comparison with stimulated emission, and (2) we eliminate as many terms as possible, using the relations between the DC terms ϕ', n', I' and P' derived in the previous section. The photon rate equation then reduces to:

$$j\omega\phi'' = G\phi'n'' \qquad 12.7\text{-}1$$

Similarly, the electron rate equation becomes:

$$j\omega n'' = I''/ev - n''/\tau_e - \phi''/\tau_p - G\phi'n'' \qquad 12.7\text{-}2$$

Equations 12.7-1 and 12.7-2 may then be combined, to yield the small-signal optical modulation amplitude P'' in terms of the corresponding current modulation term I'', as:

$$P''(\omega) = (hcI''/e\lambda_g)\,(G\phi'/\tau_p) \,/\, \{-\omega^2 + j\omega[1/\tau_e + G\phi'] + G\phi'/\tau_p\} \qquad 12.7\text{-}3$$

Because this expression is rather complicated, we now introduce a number of new parameters. Firstly, we define the modulation amplitude at zero frequency as:

$$P''(0) = hcI''/e\lambda_g \qquad 12.7\text{-}4$$

Using this, we define the complex modulation ratio as $m(\omega) = P''(\omega) / P''(0)$. Finally, we introduce a resonant frequency ω_r and a damping factor ζ, defined as:

$$\omega_r = \sqrt{(G\phi'/\tau_p)} \quad \text{and} \quad \zeta = \{1/\tau_e + \tau_p\omega^{r2}\} / 2\omega_r \qquad 12.7\text{-}5$$

With this new notation, the modulation ratio may be written as:

$$m(\omega) = \omega_r^2 \,/\, \{(\omega_r^2 - \omega^2) + 2j\zeta\omega\omega_r\} \qquad 12.7\text{-}6$$

The small-signal frequency response of a laser is therefore that of a damped simple harmonic oscillator. This result might well have been anticipated; since the rate equations amount to two coupled first-order differential equations, the transient response is effectively described by two uncoupled second-order equations, each in the form of the governing equation of an oscillator. Since $m(\omega)$ is complex, the optical modulation must lag behind the current once again. The modulus of $m(\omega)$ is of most interest, since this describes the depth of the optical modulation. This is given by:

$$|m(\omega)| = \omega_r^2 \,/\, \sqrt{\{(\omega_r^2 - \omega^2)^2 + (2\zeta\omega\omega_r)^2\}} \qquad 12.7\text{-}7$$

Figure 12.7-1 shows a logarithmic plot of $|m(\omega)|$ versus ω, for typical values of ω_r and ζ. At low angular frequencies, $|m(\omega)|$ is unity. As the frequency rises, there is a peak in the response, due to a resonance between the electron and photon populations. This is known as

a **relaxation resonance**; for low damping, it occurs at a frequency close to ω_r. Finally, at high frequencies, the response falls off roughly as $|m(\omega)| = \omega_r^2/\omega^2$. The bandwidth is therefore set almost entirely by the resonant frequency ω_r. Using some simple manipulations, it is possible to define ω_r alternatively as:

$$\omega_r = \sqrt{(G\lambda_g P'/hcv)} \qquad 12.7\text{-}8$$

The resonant frequency is therefore variable, rising as the square root of the DC power level. In order to obtain the greatest modulation bandwidth, the laser should therefore be driven as hard as possible.

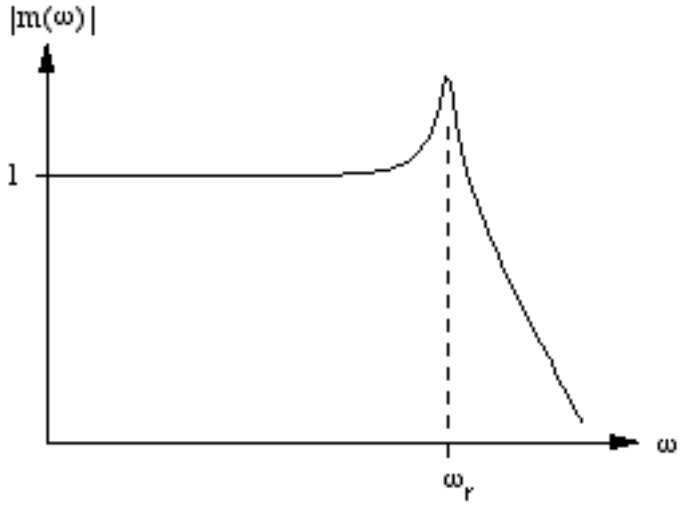


Figure 12.7-1 Small-signal frequency response of a semiconductor laser.

DESIGN EXAMPLE

We may estimate ω_r and ζ for a GaAs laser as follows. Assuming a cavity length of 250 μm, and an active area cross-section of 0.1 μm x 3 μm, a gain constant of 10^{-12} m^3/s, and an output power level of 3 mW at λ_g = 820 nm, we obtain $\omega_r = \sqrt{\{10^{-12} \times 0.82 \times 10^{-6} \times 3 \times 10^{-3} / (6.62 \times 10^{-34} \times 3 \times 10^8 \times 250 \times 0.1 \times 3 \times 10^{-18})\}} = 1.29 \times 10^{10}$ rad/sec, corresponding to a bandwidth of ≈ 2 GHz.

This is entirely typical; the modulation bandwidth of a laser is typically ≈ 100x greater than that of an LED, making the laser the preferred source for high bit-rate communications systems. For modulation at still higher speeds, an external modulator must be used. Now, we have previously estimated the photon lifetime for a 250 μm - long GaAs laser as $\tau_p \approx 2.5$ psec. Assuming a recombination time of $\tau_e \approx 1$ nsec, the damping factor is $\zeta = \{10^9 + 2.5 \times 10^{-12} \times (1.29 \times 10^{10})^2\} / (2 \times 1.29 \times 10^{10}) \approx 0.05$. This is also typical; the relaxation oscillations are underdamped at normal powers, leading to the sharp peak in the response near ω_r in Figure 12.7-1.

12.8 DBR AND DFB LASERS

Although it is blessed with the important advantage of simplicity, the Fabry-Perot semiconductor laser suffers from several undesirable characteristics. The first is that its output is often multi-line, a major drawback for optical communication systems. In this Section, we shall consider a number of solutions to the problem. The two most successful

approaches involve reflection gratings, of the type described in Chapter 10; these are known as the **distributed Bragg reflector** (or DBR) and the **distributed feedback** (or DFB) laser, respectively. Figure 12.8-1 is a comparison between the two structures.

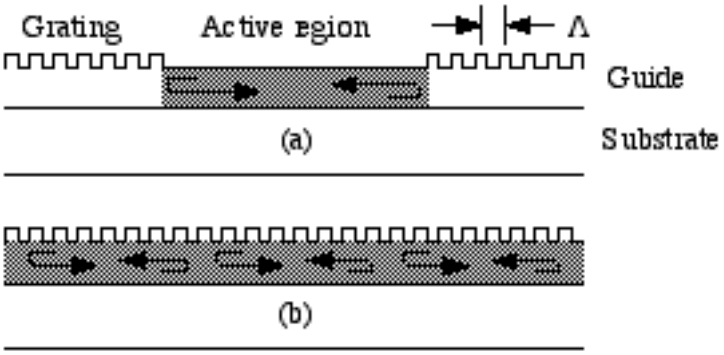


Figure 12.8-1 a) DBR and b) DFB lasers.

In the DBR laser, either or both of the cleaved end-mirrors are replaced by a corrugated reflection grating. As a result, the DBR laser has several sections: a central active region, which provides the gain, and passive grating reflectors, which provide feedback. Now, we would expect high reflectivity from the grating only for longitudinal modes whose wavelengths approximately satisfy the Bragg condition, $2n_{effv}\Lambda = \lambda_v$. When this is not satisfied, the reflectivity is quickly reduced to zero, drastically reducing the photon lifetime in the process. We have already shown that the relative importance of different longitudinal modes is governed by amplification factors F_v. Including a mode- dependent photon lifetime τ_{pv}, these might be written as $F_v = 1 / \{1/\tau_{pv} - [G(n - n_0) + b'(\lambda_v - \lambda_p)^2]\}$. Since they depend inversely on the difference between two comparable quantities, even a small increase in $1/\tau_{pv}$ can have a significant effect. Consequently, the introduction of grating feedback causes a vastly greater purification of the spectrum than the grating filter response might suggest. In fact, the mode selectivity is governed almost entirely by the curvature of the response near the Bragg wavelength, and the output of a DBR laser may consist only of a single line.

Now, the wavelength of each longitudinal mode is governed by the optical path length between the reflectors. Since the effective index in the active region will depend on the level of pumping, the positions of the longitudinal modes relative to the peak in grating reflectivity must alter with current. To obtain optimum reflectivity for one desired mode, independent of variations in drive current, a slightly more complicated structure is required. For example, Figure 12.8-2 shows a three-section DBR laser.

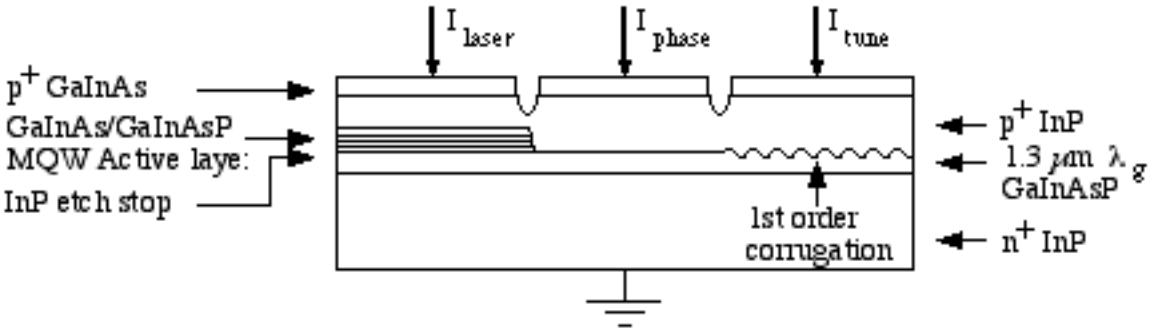


Figure 12.8-2 A three-section, continuously tunable MQW-DBR laser (after T.L.Koch et al., Elect. Lett. 24, 1431-1433 (1988)).

This has a multi-quantum well active section; feedback is provided by a cleaved mirror at the left-hand end, and a grating reflector at the right. However, in this case the Bragg wavelength may be altered by a tuning current, which varies the effective index beneath the grating through minority carrier injection. A central passive phase-shifting section (operating by the same mechanism) then allows the position of one longitudinal mode to be matched to the peak in grating reflectivity. Through an interplay between the three currents, tunable single-line operation is therefore possible.

DFB lasers (Figure 12.8-1b) achieve a similar end result, but operate by an entirely different principle. In this case, the grating and the active region coincide, and is no longer possible to consider the feedback as being provided by reflectors at either end of a cavity; instead, it must be distributed throughout the active region. To show that this may give rise to laser operation, we need a simple model. Below threshold, linear coupled mode theory is suitable. All that is required is the insertion of an effective gain parameter g into the coupled equations previously derived for a reflection grating. If this modification is made to Equations 10.10-14, for example, we obtain:

$$dA_F/dz - gA_F + j\kappa A_B = 0$$
$$dA_B/dz + \{g - j2\Delta\beta\}A_B - j\kappa A_F = 0 \qquad 12.8\text{-}1$$

We can check that the extra terms are realistic, by examining the solutions for the forward and backward mode amplitudes A_F and A_B when the coupling coefficient is zero. They grow exponentially in the $\pm z$ directions, as $\exp(\pm gz)$, which seems reasonable. When κ is non-zero, the equations may still be solved analytically. However, the solutions are hard to interpret, as they contain many complex quantities. It is therefore simpler to discuss a typical numerical solution, for the boundary conditions of unity input to one end of the device (i.e., for $A_F = 1$ on $z = 0$ and $A_B = 0$ on $z = L$). For example, Figure 12.8-3 shows logarithmic plots of reflectivity $R = A_B A_B^*$ on $z = 0$ versus detuning $\Delta\beta L$, for a coupling length of $\kappa L = 0.4$ and different values of normalised gain gL.

For $gL = 0$, the response is that of a passive grating, reaching a peak when the Bragg condition is satisfied and falling gradually to zero with a standard filter response as $|\Delta\beta L|$ increases. As gL rises, however, the shape of the response alters dramatically. The reflectivity improves overall, which is only to be expected. More interestingly, sharp peaks start to appear in the curves. These are displaced symmetrically on either side of the Bragg condition, and their characteristics are broadly analogous to those of the longitudinal resonances occurring in a Fabry-Perot cavity. They arise when the contributions that are scattered by the grating and then amplified by the gain mechanism combine with a suitable relative phase to ensure high net reflectivity. For $gL = 3$, the first-order resonances are extremely pronounced, reaching a peak of $\approx 13{,}000$, a staggering figure. In fact, for slightly higher gain, R may be infinite, as may the transmissivity T. At this point, the device may be said to be lasing, because an output may appear from both ports for a zero input.

From Figure 12.8-3, it is clear that the largest peaks occur at the first-order resonances, so these will have the lowest threshold. The spectrum of a DFB laser will therefore consist of two lines, each of slightly different wavelength. These are automatically aligned with respect to the Bragg wavelength, so no separate tuning mechanism is required to optimise their thresholds. However, the wavelength of each line will alter with drive current, due to the change in effective index, so a DFB laser is tunable over a moderate range. Although this is a considerable improvement over a Fabry-Perot laser, further advances are possible. All that is required is to generate some form of asymmetry in the grating reflectivity. This may be

done, by inserting a step in the phase of the grating at a suitable point along the cavity, a simple process for gratings fabricated by electron-beam writing. Since the required shift is one quarter of a period, such devices are known as **Λ/4-shifted DFB lasers**. The net result is that the threshold gains of the two first-order resonances are displaced from one another, so that one longitudinal mode lases preferentially and the output is single-line.

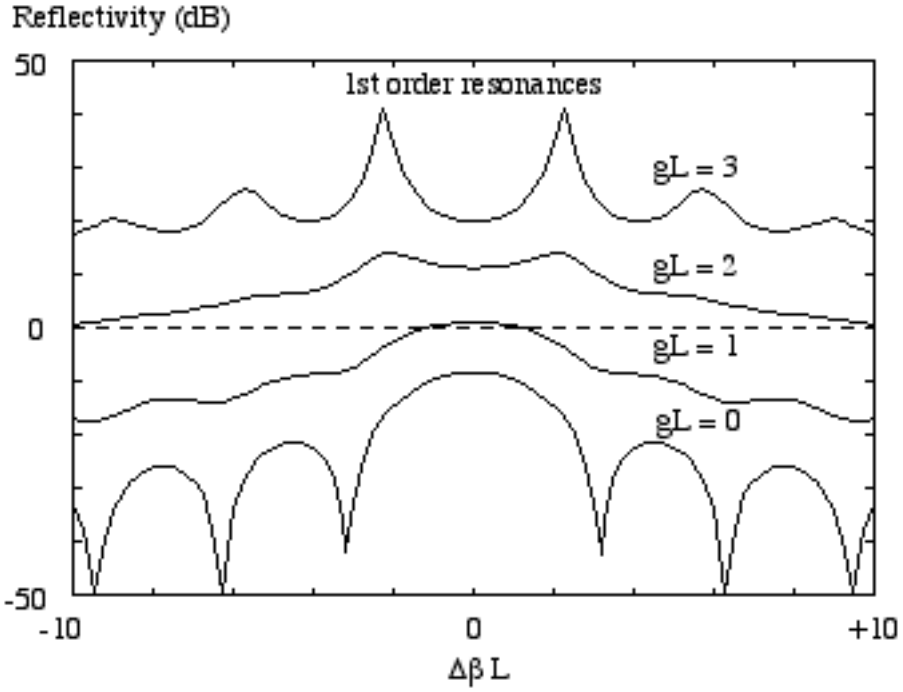


Figure 12.8-3 Reflectivity as a function of detuning ΔβL, for a reflection grating with gain.

Figure 12.8-4a shows a schematic of a typical long-wavelength ridge waveguide DFB laser. This is very similar in construction to the Fabry-Perot laser shown in Figure 12.5-4, but an extra layer of material has been inserted above the active layer. The grating is fabricated by etching a pattern of electron-beam-defined grooves into this layer, using a preferential wet chemical etch that forms self-stopping V-shaped grooves with a precise depth (giving a well-controlled coupling coefficient). Note that etching of the active layer itself is inadvisable, because most etch processes create surface damage sites that degrade the recombination efficiency, increasing the threshold.

Because it is possible to fabricate complicated grating profiles, the perturbation of the guide need not be sinusoidally varying (although it must be periodic in Λ). In general, the perturbation may be written as a Fourier series, with components of period Λ, Λ/2, Λ/3 etc. These have associated grating vectors $K_1 = 2\pi/\Lambda$, $K_2 = 2K_1$ and $K_3 = 3K_1$, respectively.

We have already seen that the fundamental component of a grating may be used to provide Bragg reflection at a wavelength $\lambda_0 = 2n_{eff}\Lambda$ (say, a near-infrared wavelength). Clearly, the second harmonic component will operate in a similar way at half this wavelength (somewhere in the visible spectrum). This is not particularly useful in itself. However, if the grating period is now doubled, the second harmonic may operate at the original IR wavelength. This is highly advantageous, because it allows the resolution of the process used to define the grating pattern to be relaxed. **Second-order gratings** are therefore common.

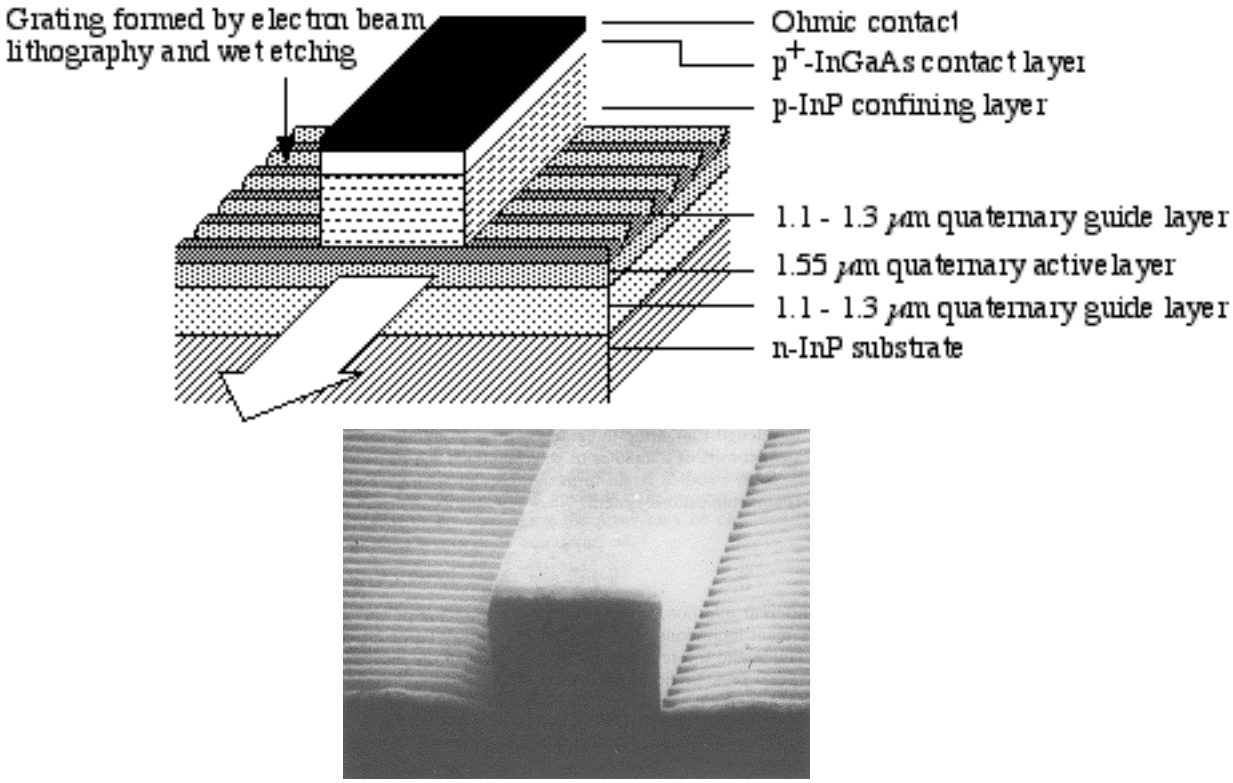


Figure 12.8-4 a) Schematic of a typical InP/InGaAsP ridge waveguide DFB laser, and b) SEM view of a similar device (photo courtesy Dr C.J.Armistead, S.T.C.Technology Ltd.)

Second-order gratings do have one additional characteristic, which is best illustrated using K-vector diagrams. Figure 12.8-5a shows the diagram for Bragg reflection using the second-harmonic grating vector K_2, which generates a backward-going diffraction order from a forward-travelling input. This is as expected. However, the fundamental grating vector K_1 may also cause diffraction, as shown in Figure 12.8-5b. Since it is half the length of K_1, the diffraction orders it generates propagate vertically upwards into the guide cover layer and downwards into the substrate. The fundamental grating therefore acts as a grating coupler, dumping energy out of the cavity. Consequently, it acts as an additional source of loss, increasing the lasing threshold. Although this is normally undesirable, many manufacturers are prepared to pay the price for the reason given above. The same output coupling mechanism is also important in surface-emitting lasers, which we shall describe shortly.

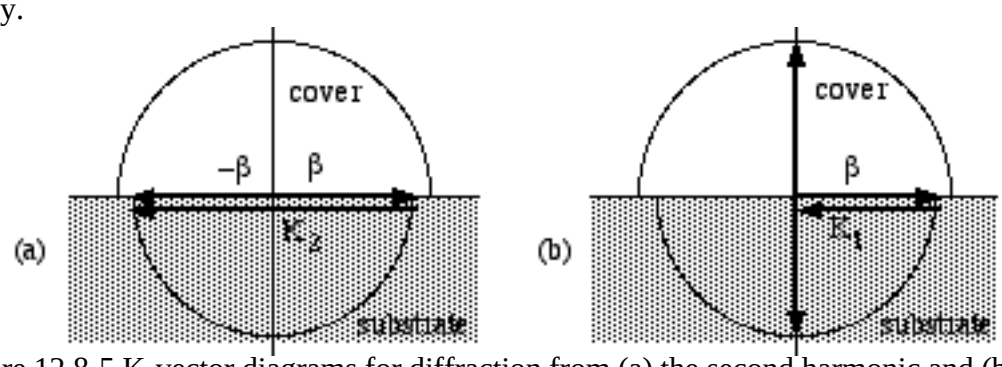


Figure 12.8-5 K-vector diagrams for diffraction from (a) the second harmonic and (b) the fundamental of a second-order grating.

All forms of grating feedback share one further advantage over Fabry-Perot lasers. Because they do not require cleaved mirrors, the substrate is not limited to the length of the laser. It is therefore simple to combine a DFB or DBR laser with other guided wave devices. Particular attention is now being paid to the integration of arrays of lasers with beam-combining components, so that their outputs may be combined in a single channel. For example, an array of DFB lasers (each arranged to emit a slightly different wavelength) might be integrated with a Y-junction tree. The output of the tree will then consist of a number of wavelengths, each of which can be modulated separately. Such a source might form the basis of a wavelength-division multiplexed (WDM) communications system.

12.9 ARRAY LASERS

We shall now consider a different question, namely how to increase the power of a semiconductor laser. Now, the main limitations of stripe-geometry Fabry-Perot lasers are their restricted active volume and the requirement that the optical output must pass through end-facets of small cross- sectional area - indeed, catastrophic mirror failure was a major problem in early devices. One obvious way to improve total output is to use several lasers, placed side-by-side on a chip, so that the active volume and output-coupling window are increased simultaneously. The problem with this arrangement is that each of the lasers will oscillate independently. As a result, it is impossible to combine their outputs into a single, high-power beam, because they must sum together with random phases.

To obtain a useful increase in power, the laser must be forced to emit light of a fixed spatial pattern. Preferably, the outputs of all emitters should be of equal amplitude and phase, and a variety of possible structures have been investigated in pursuit of this goal. For example, the earliest geometry - the **coupled waveguide array laser**, shown in Figure 12.9-1 - was based on a combination of a directional coupler and a Fabry-Perot cavity. A number of parallel laser stripes were arranged close together on a chip, so that the modal fields in adjacent stripes overlapped. It was anticipated that co-directional coupling would then introduce **phase locking**. Unfortunately, this was not the case, as we show below; however, the experience allowed better structures to be devised very quickly.

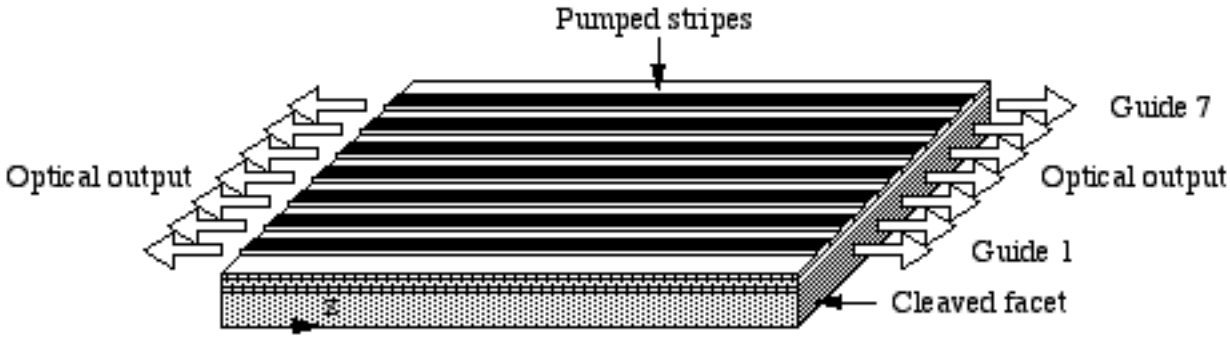


Figure 12.9-1 A seven-element co-directionally coupled ridge waveguide array laser.

To analyse the coupled waveguide array laser, we need a simple model. Once again, this may be provided by coupled mode theory. Ignoring the feedback from the mirrors for the time being, we may start by considering the related problem of wave propagation in an N-element coupled waveguide array. We shall assume there is equal gain in each stripe (defined by an effective gain g) and nearest- neighbour coupling between stripes (defined by a coupling coefficient κ). In this case, the coupled mode equations governing the change

in amplitude of the mode in each stripe with distance along the array may be obtained by a simple modification of our earlier two-guide equations, as:

$$\begin{gathered} dA_1/dz - gA_1 + j\kappa A_2 = 0 \\ dA_n/dz - gA_n + j\kappa \{A_{n+1} + A_{n-1}\} = 0 \quad \text{for } n = 2, 3 \ldots N - 1 \\ dA_N/dz - gA_N + j\kappa A_{N-1} = 0 \end{gathered}$$

12.9-1

Equation 12.9-1 shows that the two outermost guides are coupled to a single neighbour, while all the other guides are coupled to two. Now, in Chapter 10, we solved two similar equations for fixed boundary conditions. This time, we shall follow a different course and look for eigenmode solutions. These describe patterns of mode amplitude that may propagate along the array with a change in overall amplitude or phase, but not of shape. Each such pattern may therefore be written in the form:

$$A_n = a_n \exp(-j\gamma z)$$

12.9-2

Here the coefficients an describe the fixed shape of the pattern, while the eigenvalue γ (which may be complex) accounts for any change in gain or phase additional to the average propagation constant β_0 implicit in the equations. Initially, we shall assume that the array is infinite, so that the central equation in 12.9-1 is valid everywhere. Because the array now represents an infinite 'medium' (albeit with rather complicated properties), it is natural to look for wave solutions. We therefore put:

$$a_n = A \exp(-j\phi n)$$

12.9-3

where ϕ is a constant. Equation 12.9-3 describes a form of travelling wave, moving across the array and changing in phase by discrete amounts between guides. There is a strong analogy between this solution and the Bloch waves of Chapter 11; indeed, the behaviour of light in a periodic array of guides is very similar to that of electrons in an array of potential wells. Here we have approached the mathematics slightly differently, but we should not be surprised if similar conclusions are reached. Substituting this solution into Equation 12.9-1 and eliminating all common terms gives the following relation between γ and ϕ:

$$\gamma = jg + 2\kappa \cos(\phi)$$

12.9-4

Ignoring the gain term, Equation 12.9-4 shows that the eigenvalue γ may only take a restricted range of values (so that $-2\kappa \leq \gamma \leq 2\kappa$), varying cosinusoidally with ϕ over the range. This should be compared with the $\mathbb{E}$ - k relation predicted by the Krönig-Penney model in Chapter 11.

We now reintroduce the finite lateral extent of the array, by assuming that the upper and lower equations in 12.9-1 will effectively be satisfied if we assume coupling to two ficticious guides (at n = 0 and n = N + 1), which support modes of zero amplitude. This may be done by taking $a_n = A' \sin(\phi n)$, i.e. by assuming a standing-wave solution. Applying the boundary conditions $a_0 = a_{N+1} = 0$, we then find that ϕ is restricted to the following set of discrete values:

$$\phi_\nu = \nu\pi / (N+1)$$

12.9-5

where $\nu = 1, 2 \ldots N$ is the mode number. The complete solution to Equation 12.9-2 is therefore defined by mode amplitudes $a_{\nu n}$ and propagation terms γ_ν, given by:

$$a_{\nu n} = A \sin[\nu n\pi / (N+1)] \quad ; \quad \gamma_\nu = jg + 2\kappa \cos[\nu\pi / (N+1)]$$

12.9-6

Now, the terms $a_{\nu n}$ are the amplitudes of the modes in each guide. If each individual mode is defined by a transverse field $E_i(x, y)$, the total field pattern of the ν^{th} eigenmode must therefore be:

$$E_\nu(x, y) = \sum_{i=1}^{N} a_{\nu i} E_i(x, y)$$

12.9-7

Figure 12.9-2a shows a one-dimensional plot of the lowest-order transverse mode ($\nu = 1$) for a seven- guide array. Within each guide, the pattern is like that of the individual guide mode, but there is an overall sinusoidal envelope defined by Equation 12.9-6. Similarly, Figure 12.9-2b shows the second- order mode ($\nu = 2$). This is similar, but the spatial frequency of the envelope is now doubled. The trend continues as ν increases, so that the envelope of the highest-order mode has the greatest spatial frequency, with adjacent mode amplitudes of opposite sign.

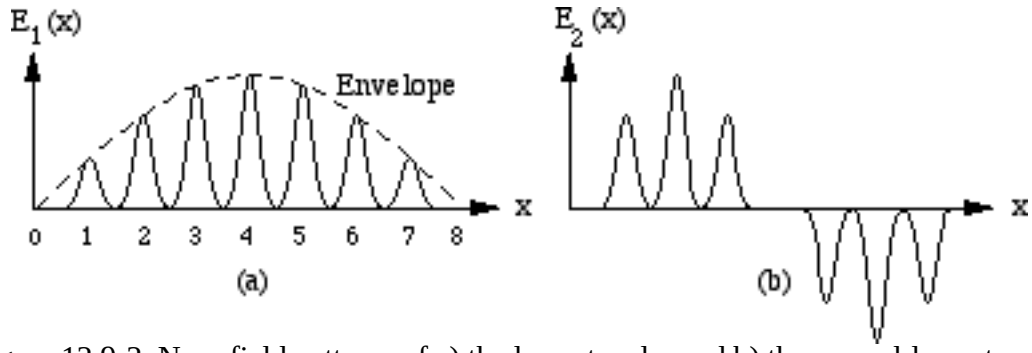


Figure 12.9-2 Near-field patterns of a) the lowest-order and b) the second-lowest order transverse mode of a seven-element coupled waveguide array.

We now return to the provision of feedback by the end-facets. If their reflectivity is uniform across the array, each eigenmode will be reflected without any change in shape, i.e. without mode conversion. The eigenmodes found above are therefore identical to the transverse modes of the waveguide array laser itself, since they may travel a round-trip through the cavity unaltered. If the laser can be forced to oscillate on just one of these modes (the lowest-order is the most uniform, and hence the most useful), the required phase-locking will have been introduced, because the output from each stripe will then have a fixed phase relationship with its neighbours.

To investigate the mode selectivity of the structure, we must examine the conditions for lasing. The total propagation constant of the ν^{th} transverse mode is $\beta_0 + \gamma_\nu$, including the average term β_0. Only $\mathrm{Re}(\gamma_\nu)$ will affect phase variations, so the longitudinal resonance condition is therefore:

$$2\Lambda\{\beta_0 + \mathrm{Re}(\gamma_\nu)\} = 2\mu\pi$$

12.9-8

where μ is an integer and L is the cavity length. Since $\mathrm{Re}(\gamma_\nu)$ is different for each transverse mode, we may deduce that each will have a different spectrum of longitudinal modes. Similarly, only $\mathrm{Im}(\gamma_\nu)$ governs the overall gain of each mode, so the condition for unity round-trip gain is:

$$R_1 R_2 \exp\{2L\ \mathrm{Im}(\gamma_\nu)\} = 1$$

12.9-9

where R_1 and R_2 are the amplitude reflectivities of the mirrors. Examining Equation 12.9-6, however, we can see that $\mathrm{Im}(\gamma_\nu) = g$ for all ν. Consequently, the threshold gain must be the same for each transverse mode, so co-directional coupling has actually introduced no selectivity at all. We would therefore expect the laser to oscillate on a random mixture of all the possible transverse modes.

In practice, early array lasers tended to lase only on the highest-order mode. The explanation lay in the fact that only the active region of each stripe is pumped, so that there is very little gain between the stripes. An accurate calculation shows that the coupling coefficient is then complex, so that $\mathrm{Im}(\gamma_\nu)$ is actually different for each mode, being greatest for the highest-order one. This is physically realistic; this particular mode has a field that changes in sign from guide-to-guide, falling to zero in between, so it is unaffected by interstripe loss. However, being the least uniform mode, it also is the least usable. Despite some attempts to introduce selectivity by separately pumping each stripe or using non-uniform stripe spacings, directionally coupled arrays have therefore largely been abandoned. Nevertheless, high outputs were achieved; for example, 2.6 W CW was obtained from a 40-stripe array.

A more effective mechanism is required force single-transverse-mode operation. The obvious solution is to introduce a mode-dependent loss, which reduces the effective gain of undesired modes while allowing desired ones to travel up and down the cavity unaffected. Now, in Chapter 9 it was shown that a single-mode Y-junction will radiate any anti-symmetric component that is injected at its forked end. Consequently, Y-junctions can be used to construct a filter that discriminates against transverse field patterns of high spatial frequency (which have fields of opposite sign in adjacent stripes). Figure 12.9-3 shows a device based on such a filter, the **Y-junction-coupled array laser**. At either end of the device, there is an array of uncoupled single-moded laser stripes. In this example, there are seven stripes in the left-hand array, and eight in the right. The two arrays are linked together in the central section by a network of back-to-back Y-junctions, leaving two unconnected stubs on the left. All sections of guide are equally pumped, and the structure is contained in a Fabry-Perot cavity.

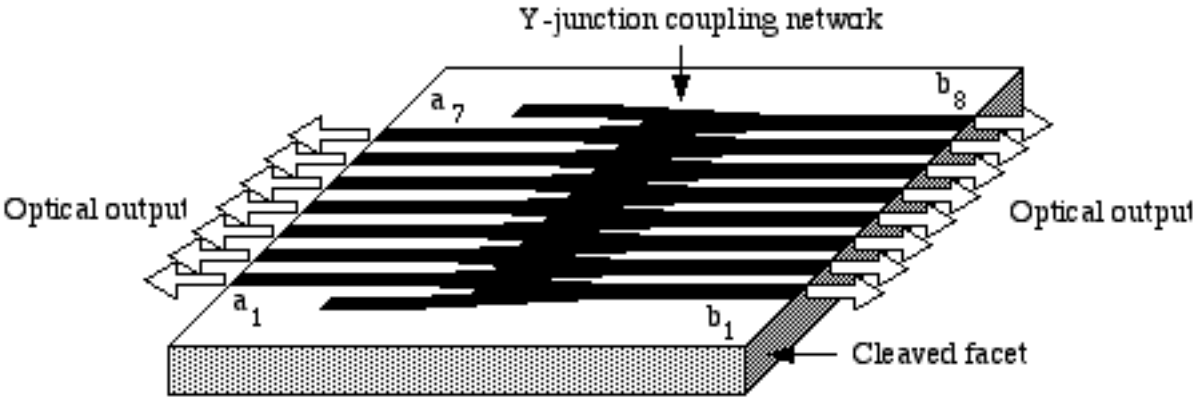


Figure 12.9-3 A Y-junction-coupled semiconductor array laser.

To analyse the structure, we assume for generality that there are N stripes at the left-hand end of the array, and N+1 at the right. Once more, we assume that each guide has a total length L, an amplitude gain g and a propagation constant β_0. Similarly, the end-mirror amplitude reflectivities are again R_1 and R_2. The Y-junctions act as linear beam splitters or combiners, as described in Chapter 9. We proceed by defining the amplitudes of the modes in the individual stripes at the left-hand end of the cavity as a_1 ... a_N at the start of a round

trip. It is then simple to show that the amplitudes b_1 ... b_{N+1} before reflection at the right-hand facet are:

$$b_n = 1/2\{a_n + a_{n-1}\} \exp(-j\beta_0 L) \exp(gL)$$

12.9-10

Here once again we have introduced two ficticious amplitudes a_0 and a_{N+1} (those of the modes in the stubs), which are both zero. After reflection at the right-hand facet, propagation back to the left-hand end of the cavity, and further reflection, the new mode amplitudes a_1' ... a_N' are:

$$a_n' = 1/4\{a_{n+1} + 2a_n + a_{n-1}\}R_1R_2 \exp(-j2\beta_0 L) \exp(2gL)$$

12.9-11

Now, if the original amplitudes a_1 ... a_N represent an eigenmode, they should be related to the new amplitudes a_1 ... a_N' by a constant factor (otherwise the pattern will have changed in a round-trip). We may thus put $a_n' = \delta a_n$, where δ is a constant. This allows us to solve Equation 12.9-11 using our previous methods. Initially, we ignore the finite extent of the array, and assume a solution as a kind of travelling wave, writing an = A exp(-jϕn). We then obtain:

$$\delta = 1/2\{1 + \cos(\phi)\} R_1R_2 \exp(-j2\beta_0 L) \exp(2gL)$$

12.9-12

However, since the array is actually finite, we must satisfy the boundary conditions $a_0 = a_{N+1} = 0$. The solution must therefore again be a standing-wave pattern, so that $a_n = A' \sin(\phi n)$. Since it is clear that the analysis will give very similar results to those obtained for the co-directionally-coupled array, we may immediately deduce that ϕ is again restricted to the set of values given in Equation 12.9-5, and that the overall shapes of the modes at the left-hand-end of the array are as specified in Equation 12.9-6.

Each mode will lase when its round-trip gain is unity, i.e. when $\delta = 1$. From Equation 12.9-12, we can see that the longitudinal resonance condition for all transverse modes is now the same, namely $2\beta_0 L = 2\mu\pi$. More interestingly, the threshold gain g_ν is now different for each, being given by:

$$g_\nu = 1/2L \log_e(1/f_\nu R_1R_2)$$

12.9-13

Here the factor f_ν (the transmission of the Y-junction network for the ν^{th} mode) is:

$$f_\nu = 1/2\{1 + \cos[\nu\pi / (N+1)])\}$$

12.9-14

From this, we can see that the highest transmission efficiency (slightly less than unity) is obtained for the lowest-order transverse mode, which must therefore have the lowest lasing threshold. In contrast, the transmission of the highest-order mode is close to zero, so its threshold will be very high. The Y-junction network has therefore introduced the desired selectivity in favour of a single transverse mode with a useful spatial distribution. Practical Y-junction-coupled lasers have excellent stability, showing little change in their near- or far-field patterns over the entire range of operation, and their output powers are typically of the order of 400 mW CW at present.

12.10 SURFACE-EMITTING LASERS

In the future, it may be desirable to arrange for the emission from a laser chip to be taken from its surface, rather than an end-facet. This would allow the fabrication of two-

dimensional arrays of lasers, which might have applications in optical processors. At the very least, it would enable an entire wafer of lasers to be tested for failures before individual devices are separated by cleaving. A variety of **surface-emitting semiconductor laser** geometries are currently being investigated. These are all still based on single-mode waveguides, but there are two major distinctions between types.

Firstly, the laser cavity may be in the conventional orientation, parallel to the surface. Feedback may be obtained without cleaving, by using distributed Bragg reflectors. If these are based on second-order gratings, diffraction from the grating fundamental can be used to obtain output coupling normal to the surface, as described earlier (although unnecessary cladding layers over the grating are generally removed, to minimise absorption). For example, Figure 12.10-1 shows a **Y-junction-coupled, surface-emitting DBR array laser**. A laser of this type has already been demonstrated with a 10 x 10 array of emitters occupying an area of 60 μm x 5 mm, with a total output power of more than 1 W.

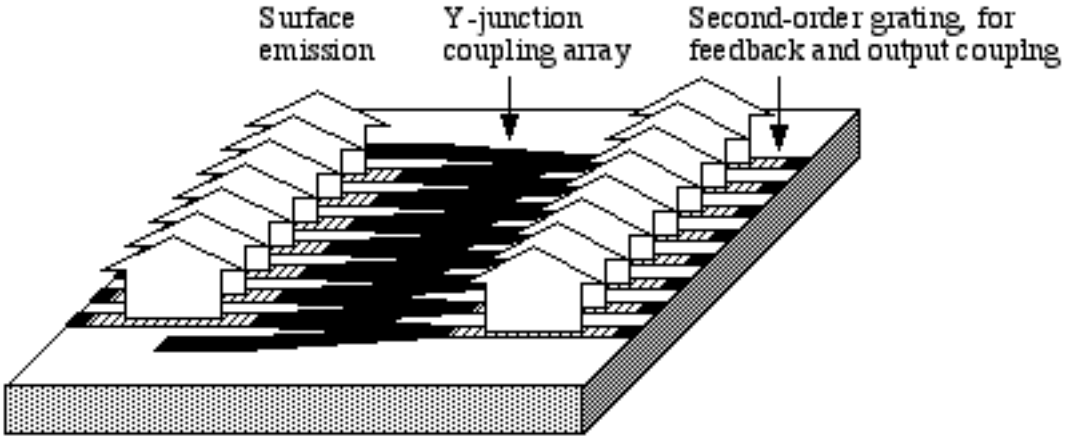


Figure 12.10-1 A Y-junction-coupled, surface-emitting DBR array laser.

Alternatively, the cavity may be normal to the substrate surface. In this case, the device is known as a **vertical cavity surface-emitting laser**. Figure 12.10-2 shows a typical example, a circular buried heterostructure GaAs/GaAlAs laser. Although the double heterostructure is still formed from layers parallel to the surface, the cavity is a roughly cylindrical pillar approximately 6 μm in diameter and 7 μm long with its axis normal to the surface. The current is injected from a ring electrode into the lower end of the pillar, which is surrounded by lateral current-blocking layers. Feedback is provided by a gold/silica mirror at the bottom of the cavity, and by a silica/titania multilayer stack deposited in a deep well etched right through the substrate at the top. The optical output is taken from the upper mirror.

More recent vertical cavity lasers have used distributed feedback or distributed Bragg reflection, based on epitaxial multilayers, thus eliminating the requirement for external mirrors. Due to the short length of the cavity, the output consists only of a single longitudinal mode, a feature of all such lasers. The particular laser shown had a threshold current of 6 mA, and an output in excess of 1 mW. Although all surface-emitting lasers are at a very early stage of development, they represent a promising new technology, which may have many applications outside communications.

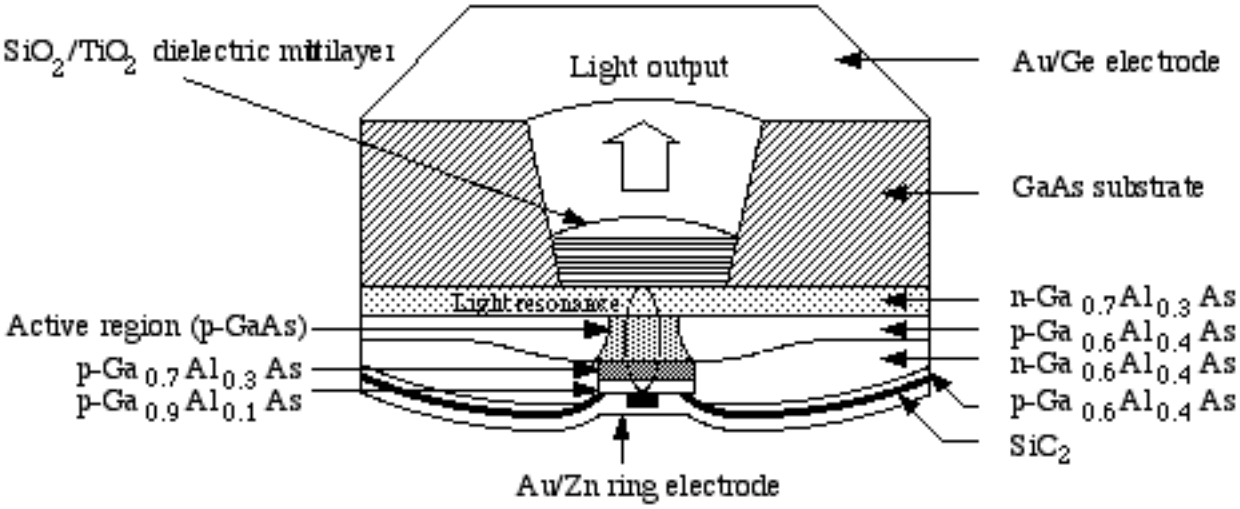


Figure 12.10-2 Cross-section of a GaAs/GaAlAs circular buried heterostructure vertical cavity surface emitting laser (after S.Kinoshita et al., IEEE J. Quant. Elect. QE-23, 882-888 (1987); © IEEE 1987)

12.11 TRAVELLING-WAVE LASER AMPLIFIERS

Besides acting as a versatile source, the semiconductor laser can provide other functions. For example, if the end-facets are suitably coated so that they are anti-reflecting, laser operation will be suppressed due to the lack of feedback. However, optical gain will still be available in the cavity, provided the electron density is driven above threshold. In this case, the device may act as a **travelling-wave laser amplifier** (TWLA). Figure 12.11-1 shows the principle, and a typical application.

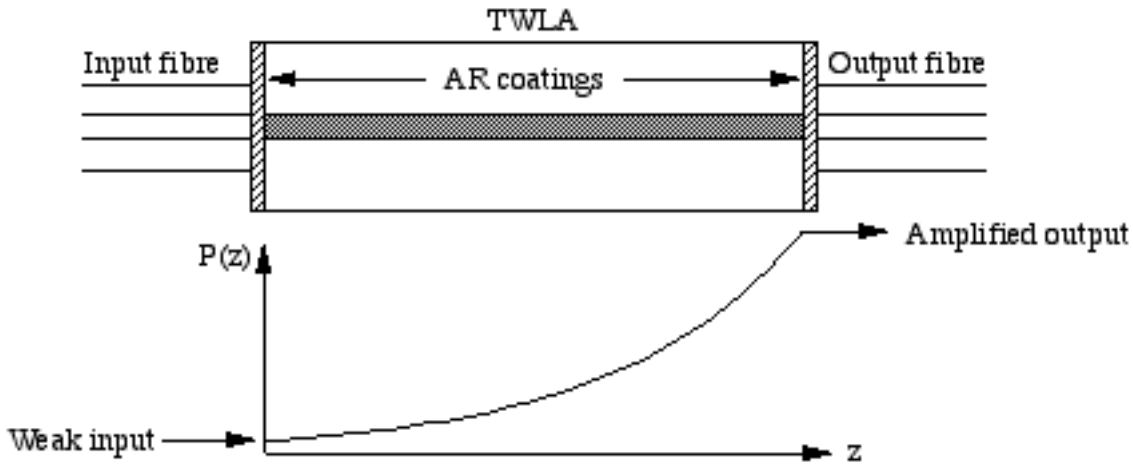


Figure 12.11-1 Schematic of a travelling-wave semiconductor laser amplifier.

The TWLA is connected between two fibres, as an in-line amplifier. A relatively weak optical input will grow exponentially as it passes through the TWLA, emerging as a much stronger amplified signal. The maximum possible input is limited, because the gain saturates at high optical power, but this level is quite sufficient for use in typical systems. Recent devices based on double-channel planar buried heterostructure lasers show little polarization sensitivity (1 dB variation in gain between orthogonal polarizations), gains of ≈ 30 dB

(although this figure is reduced by coupling losses), and bandwidths of ≈ 75 GHz. Unfortunately, the amplification is not noiseless, due to the presence of spontaneous emission. This may be minimised by passing the output through a narrowband filter, which blocks the majority of the wide-band spontaneous emission noise, while passing the desired signal. Figure 12.11-2 shows a fully packaged travelling wave laser amplifier with fibre pigtails.

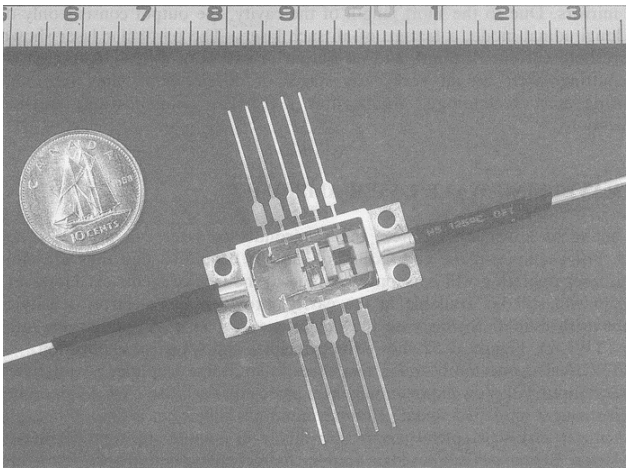


Figure 12.11-2 A travelling-wave laser amplifier (photograph courtesy *Telesis*, a publication of Bell Northern Research Ltd.).

Travelling wave laser amplifiers can have applications other than as simple repeaters. For example, Figure 12.11-3 shows an amplifying switch based on a passive coupler combined with two TWLAs. The coupler acts as a splitter, which merely directs the input to the two amplifiers. Since the output of each TWLA may be switched off by removing the bias current, the whole network acts effectively as a two-way switch. The isolation achieved in the off-state is good - typically 30 dB for a 200 μm long device - while the gain provided in the on-state is easily sufficient to make up for coupling losses. Fully integrated switches based on similar principles appear an attractive possibility for the future.

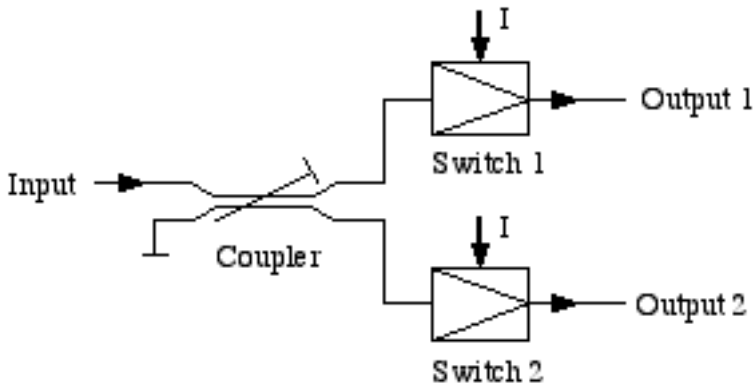


Figure 12.11-3 An amplifying switch based on travelling wave laser amplifiers (after M.J.O'Mahony, IEEE J. Lightwave Tech. LT-6, 531-544 (1988); © IEEE 1988).

PROBLEMS

12.1 A silicon p-n junction diode has doping levels of $N_a = 10^{24}$ m^{-3} and $N_d = 10^{23}$ m^{-3}. Estimate the electric field across the depletion layer, under 5 V reverse bias. Assuming that the mobilities of electrons and holes in Si are $\mu_e = 0.135$ m^2/Vs and $\mu_h = 0.048$ m^2/Vs, respectively, estimate their transit times across the depletion layer. Discuss the implications for photodiodes.
[20.4 x 10^6 V/m; 0.1 psec; 0.3 psec]

12.2 A silicon p-n$^+$ diode has a cross-sectional area of 1 mm^2, and doping levels of $N_a = 10^{21}$ m^{-3} and $N_d = 10^{24}$ m^{-3}. Assuming that the distance L_p from the depletion layer to the p-side contact is 0.1 μm, and that the corresponding distance L_n is » L_p, calculate the current through the diode under a forward voltage of 0.5 V. The hole and electron diffusion coefficients may be found using the **Einstein relation** $D_{e,h} = \mu_{e,h}kT/e$. If the minority carrier lifetimes are ≈ 1 μs, are your assumptions valid?
[284 mA]

12.3 Compare the performance of InP/InGaAsP phase modulators based on a) the electro-optic effect and b) minority carrier injection. For the former, you may assume that $r_{41} = -1.4 \times 10^{-12}$ m/V is the relevant electro-optic coefficient, that the overlap factor Γ is 0.5, and that the breakdown field is 3 x 10^7 V/m. For the latter you may assume that $m_e^*/m = 0.077$, $m_h^*/m = 0.64$, and that an upper limit on the injected carrier density is $\Delta N = 10^{24}$ m^{-3}. The refractive index of InP is 3.2, and the operating wavelength is $\lambda_0 = 1.5$ μm.

12.4 A $G_{a0.47}In_{0.53}As$ PIN photodiode, with a cross-sectional area of 0.1 mm^2, produces a photocurrent of 0.996 μA when illuminated by a beam of 1.3 μm wavelength and 1 mW/cm^2 intensity. Calculate a) the quantum efficiency, and b) the photocurrent at 1.5 μm and 1.7 μm wavelength.
[0.95; 1.15 μA; 0]

12.5 A surface-emitting GaAs LED, emitting at ≈ 820 nm, launches 50 μW of optical power into a step-index fibre when driven by a current of 50 mA. Assuming that the fibre has a NA of 0.3, estimate the internal efficiency of the LED. If the minority carrier recombination time is 10 nsec, what is the radiative recombination time? Estimate the 3 dB modulation bandwidth of the LED.
[54%; 18.4 nsec; 27.6 MHz]

12.6. Show that the effective power gain experienced by a guided mode with normalised transverse electric field E due to the presence of a non-uniform power gain distribution g_p may be written as $g_{peff} = < g_p , |E|^2 >$. Discuss the way in which the double heterostructure maximises g_{peff}.

12.7. Show that, if a non-unity confinement factor and non-zero losses are included in the calculation, the power obtained from a semiconductor laser in the lasing regime alters to $P = \eta hc\{I - I_t\}/e\lambda_g$, where:

$$\eta = \{1/L \log_e(1/R_1R_2)\} / \{\Gamma\alpha_{pcore} + (1 - \Gamma)\alpha_{pclad} + 1/L \log_e(1/R_1R_2)\}.$$

The table below shows measurements of the power out of one facet of a GaAs/ GaAlAs laser at different currents. The active volume is 250 x 0.1 x 3 μm, and

emission is at 820 nm. Estimate the threshold current density, and the average power absorption coefficient in the cavity.

I	P
20 mA	15 μW
30 mA	500 μW
35 mA	2 mW
37.5 mA	2.75 mW

12.8 Assuming that the drive current of a semiconductor laser is in the form of a DC bias current I' plus a small time-varying modulation I"(t), and that the optical output power is the sum of a constant term P' and a similar time-varying term P"(t), show that the differential equation governing P"(t) is:

$$d^2P''/dt^2 + 2\zeta\omega_r\, dP''/dt + \omega_r^2\, P'' = \{h c\omega_r^2/e\lambda_g\}I'',$$

where the resonant frequency and damping factor are

$$\omega_r = \sqrt{(G\lambda_g P'/hcv)} \quad \text{and} \quad \zeta = \{1/\tau_e + \tau_p\omega_r^2\} / 2\omega_r.$$

12.9 Solve the differential equation in Question 12.8 for the case of a step change in the modulating current from I" = 0 to I" = i" at t = 0, assuming ζ is small. Sketch the response for $\zeta = 0.05$.

12.10 A Y-junction-coupled GaAs/GaAlAs array laser has nine lasing stripes at one end of the cavity, and ten at the other. Calculate the lasing threshold for all the transverse modes, assuming a cavity length of 500 μm. Sketch the near-field patterns of the lowest- and highest-order transverse modes at either end of the cavity. How would you calculate the far-field patterns?

SUGGESTIONS FOR FURTHER READING

Botez D., Ettenberg M. "Comparison of surface- and edge-emitting LEDs for use in fiber-optical communications" IEEE Trans. Electron Devices ED-26, 1230-1238 (1979)

Burrus C.A., Miller B.I. "Small-area double-heterostructure aluminium-gallium arsenide electroluminescent diode sources for optical fiber transmission lines" Opt. Comm. 4, 307-369 (1971)

Carenco A., Menigaux L., Linh N.T. "InP electro-optic directional coupler" Appl. Phys. Lett. 40, 653-655 (1982)

Casey H.C., Panish M.B. "Heterostructure lasers" Academic Press, New York (1978)

Casey H.C., Somekh S., Ilegems M. "Room-temperature operation of low-threshold separate- confinement heterostructure injection laser with distributed feedback" Appl. Phys. Lett. 27, 142-144 (1975)

Chen K.-L., Wang S. "Single-lobe symmetric coupled laser arrays" Elect. Lett. 21, 347-349 (1985)

Gillessen K., Schairer W. "Light emitting diodes - an introduction", Prentice-Hall International, Englewood Cliffs (1987).

Dutta N.K., Cella T., Piccirilli A.B., Brown R.L. "Tunable single wavelength lasers" Proc. SPIE 800, 56-62 (1987)

Evans G.A. et al. "Coherent, monolithic two-dimensional (10 x 10) laser arrays using grating surface emission" Appl. Phys. Lett. 53, 2123-2125 (1988)

Holonyak N., Kolbas R.M., Dupuis R.D., Dapkus P.D. "Quantum well heterostructure lasers" IEEE J. Quant. Elect. QE-16, 170-185 (1980)

Inoue H., Hiruma K., Ishida K., Sato H., Matsumura H. "Switching characteristics of GaAs directional coupler optical switches" Appl. Opt. 25, 1484-1489 (1986)

Kapon E., Katz J., Yariv A. "Supermode analysis of phase-locked arrays of semiconductor lasers" Opt. Lett. 10, 125-127 (1984)

Kinoshita S., Iga K. "Circular buried heterostructure (CBH) GaAlAs/GaAs surface emitting lasers" IEEE J. Quant. Elect. QE-23, 882-888 (1987)

Mikami O., Nakagome H. "InGaAsP/InP optical waveguide switch operated by carrier-induced change in the refractive index" Opt. Quant. Elect. 17, 449-455 (1985)

Moss R.H., Ritchie S. "The development of gallium indium arsenide for photodetectors" Br. Telecom J. 1, 7-22 (1983)

Nelson A.W., Spurdens P.C., Close S., Walling R.H., Moss R.H., Wong S., Harding M.J., Cooper D.M., Devlin W.J., Robertson M.J. "The role of MOVPE in the manufacture of high performance InP based optoelectronic devices" J. Crystal Growth 93, 792-802 (1988)

Okuda H., Hirayama Y., Furuyama H., Kinoshita J.-I., Nakamura M. "Five-wavelength integrated DFB laser arrays with quarter-wave-shifted structures" IEEE J. Quant. Elect. QE-23, 843-847 (1987)

O'Mahony M.J. "Semiconductor laser optical amplifiers for use in future fiber systems" IEEE J. Lightwave Tech. LT-6, 531-544 (1988)

Scifres D.R., Lindstrom C., Burnham R.D., Streifer W., Paoli T.L. "Phase locked (GaAl)As laser diode emitting 2.6W CW from a single mirror" Elect. Lett. 19, 169-171 (1983)

Stillman G.E., Cook L.W., Tabatabaie N., Bulman G.E., Robbins V.M. "InGaAsP photodiodes" IEEE Trans. Electron Devices ED-30, 364-381 (1983)

Streifer W., Cross P.S., Welch D.F., Scifres D.R. "Analysis of a Y-junction semiconductor laser array" Appl. Phys. Lett. 49, 58-60 (1986)

Thompson G.H.B. "Physics of semiconductor laser devices" John Wiley and Sons, New York (1980).

Wada O., Yamakoshi S., Hamaguchi H., Sanada T., Nishitani Y., Sakurai T. "Performance and reliability of high-radiance InGaAsP/InP DH LEDs operating in the 1.15 - 1.5 μm wavelength region" IEEE J. Quant. Elect. QE-18, 368-373 (1982)

Welch D.F., Cross P., Scifres D., Streifer W., Burnham R.D. "In-phase emission from index-guided laser array up to 400 mW" Elect. Lett. 22, 293-294 (1986)
Yariv A. "Optical Electronics", Holt, Rinehart and Winston, New York (1985); Chs. 11, 15.